

### Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung

Thaller, Manfred (Ed.)

Veröffentlichungsversion / Published Version  
Sammelwerk / collection

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Thaller, M. (Hrsg.). (1986). *Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung* (Historisch-Sozialwissenschaftliche Forschungen : quantitative sozialwissenschaftliche Analysen von historischen und prozeß-produzierten Daten, 20). St. Katharinen: Scripta Mercaturae Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-340751>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

GES THAD 1

**HSF**

**HISTORISCH - SOZIALWISSENSCHAFTLICHE FORSCHUNGEN**

**Quantitative sozialwissenschaftliche Analysen  
von historischen und prozeß-produzierten Daten**

**Herausgegeben von**

**Heinrich Best, Wolfgang Bick, Paul J. Müller,  
Herbert Reinke, Wilhelm H. Schröder**

**Zentrum für historische Sozialforschung**

**Band 20**



**Manfred Thaller (Hrsg.)**

**DATENBANKEN UND DATENVERWALTUNSSYSTEME  
ALS WERKZEUGE HISTORISCHER FORSCHUNG**

**SCRIPTA MERCATURAE VERLAG**



Die Druckvorlagen zu diesem Band wurden an der Rechenanlage der Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, mit Hilfe des Satzprogrammes  $\text{T}_{\text{E}}\text{X}$  erstellt.

**Manfred Thaller (Hrsg.)**

**Datenbanken und Datenverwaltungssysteme  
als Werkzeuge historischer Forschung**

St. Katharinen 1986

ISBN 3 - 922661 - 30 - 0

Copyright: Scripta Mercaturae Verlag

D - 6551 St. Katharinen

ISBN 3 - 922661 - 30 - 0

## Inhaltsverzeichnis

	Vorwort	5
--	---------	---

Manfred Thaller	Vorschlag für einen internationalen Workshop über internationale Quellenbanken	9
Jörn Sieglerschmidt	Überlegungen zu einer Datenbank "Historische Demographie"	31
Christhard Schrenk	Urbarauswertung mit Hilfe des Computers	43
Christian Pfister, Werner u. Gabriela Schwarz-Zanetti	Ein Programm- und Methodenpaket zur Rekonstruktion von Klimaverhältnissen seit dem Hochmittelalter	75
Gerhard Jaritz, Albert Müller	Historia Vaga. Ein computergestütztes Projekt zur Migrationsgeschichte des 15. und 16. Jahrhunderts	93
Herbert Reinke	Datenbeschreibung und Datendokumentation in der historischen Sozialforschung	125
Kevin Schurer	Historical Databases and the Researcher	145
Kevin Schurer, Richard Wall	Computing the History of the Family: A Question of Standards	159
Konrad H. Jarausch	Some Reflections on Coding	175
Helga Schultz	Probleme sozialökonomischer Klassifikation	179
Werner Braatz, Virginia Crane	A Micro Based Data Base for Undergraduates	187
Ingo Kropac	Von der Quelle zum Datensatz	199
Don Karl Rowney	The Microcomputer in Historical Research: Accessing Commercial Databases	217
	Anschriften der Verfasser	233



## Vorwort

“Datenbanken” und “Datenverwaltungssysteme” sind in der Informatik sehr präzise bestimmte – wenn auch nicht völlig unumstrittene – Begriffe. In der historischen Sozialforschung wird dieser Begriff sehr viel weiter verstanden: wenn wir als Historiker von einer Datenbank sprechen, wollen wir damit meistens die Tatsache ausdrücken, daß bei einem bestimmten Forschungsprojekt, bei einer bestimmten Institution, große Mengen von Daten in maschinenlesbarer Form lagern, die für eine Vielzahl durch ein übergreifendes Thema zusammengehaltener Forschungsfragen verwendet werden können. Dieses weiche Verständnis des Begriffes der “Datenbank” war auch bei der überwiegenden Zahl der Teilnehmer an einem Workshop vorherrschend, den das Max-Planck-Institut für Geschichte in Göttingen vom 15. bis zum 18. Juli 1985 unter dem Titel “Internationaler Workshop über die Schaffung, Verbindung und Nutzung großer interdisziplinärer Quellenbanken in den historischen Wissenschaften” abhielt. Die Veranstaltung war unternommen worden, um abzuklären, welche praktischen Möglichkeiten es gäbe, um die Bedingungen für die Nutzung von “Datenbanken” im weitesten Sinne im Bereich historischer Forschung zu verbessern. Insbesondere galt das Interesse dabei der Möglichkeit, durch Standardisierung oder andere Maßnahmen historisches Datenmaterial zwischen einzelnen Institutionen leichter austauschbar zu machen.

Während dieser Veranstaltung zeigte sich die mangelnde Präzision des eben erwähnten Begriffes recht deutlich: es bestand weitgehendes Einverständnis, daß Datenbanken, Datenverwaltungssysteme, Datenarchive und ganz allgemein jede Bestrebung, die vorhandenen historisch relevanten maschinenlesbaren Informationen leichter zugänglich zu machen, von hoher potentieller Bedeutung für den Rechneinsatz in den historischen Disziplinen seien. In den Diskussionen, was notwendig sei um dieses Potential zu nützen, zeigten sich dann freilich stark unterschiedliche Auffassungen, die im wesentlichen wohl alle auf das unterschiedliche Verständnis einer “historischen Datenbank” zurückgingen.

Diese Meinungsverschiedenheiten waren keineswegs unfruchtbar, sie haben im Gegenteil dazu geführt, daß eine ganze Reihe der Teilnehmer auf Möglichkeiten aufmerksam wurde, die über ihr ursprüngliches Verständnis hinausgingen. Deshalb möchte der Herausgeber mit diesem Aufsatzband versuchen, einen Überblick über die derzeitigen Vorstellungen vom Sinn und der Existenzberechtigung historischer Datenbanken zu geben. Dennoch ist die vorliegende Sammlung keine Veröffentlichung von Proceedings: die Göttinger Veranstaltung suchte bewußt das völlig interdisziplinäre Gespräch und brach-

te daher Projekte und Institutionen, die große Textkorpora verwalten, archäologische Einrichtungen betreuen oder historisch-sozialwissenschaftliche Forschung betreiben, an einen Tisch. Dies war für das Gespräch sehr fruchtbar – der Herausgeber mußte jedoch befürchten, daß eine Sammlung aller schriftlichen Beiträge als Buch einen zu inkohärenten Charakter gehabt hätte.

Der vorliegende Band wendet sich daher ganz bewußt an die Vertreter einer historischen Sozialwissenschaft und versammelt nur die hier einschlägigen – um zwei weitere ergänzten<sup>1</sup> – Workshopbeiträge.

Die Anordnung des Materials spiegelt folgenden Gedankengang wieder: der Herausgeber möchte zunächst selbst nochmals die Notwendigkeit einer verbesserten Zusammenarbeit zwischen den verschiedenen den Rechner nützenden historischen Forschungsprojekten und Institutionen unterstreichen. Danach sollen die Beiträge von Sieglerschmidt, Schrenk, Pfister/Schwarz-Zanetti und Jaritz/Müller einen Überblick über die Bandbreite innerhalb der Forschung derzeit angelegter, betriebener und konzipierter Datenbanken geben. Während das Schwergewicht hier eindeutig auf den datenerzeugenden, inhaltlich arbeitenden Forschungsprojekten liegt – wenn auch in jedem einzelnen Fall Überlegungen über die Nutzbarkeit durch Dritte angestellt wurden – diskutieren Reinke und Schurer/Wall anschließend die Vorbedingungen für einen erfolgreichen Datenaustausch zwischen derartigen Projekten aus der Warte damit über das einzelne inhaltliche Forschungsinteresse hinaus befaßter Forschungsunternehmen.

Spätestens zu diesem Zeitpunkt sollte es dem Leser bewußt werden, daß es bei der Diskussion über alle Maßnahmen, die der Verbesserung der Nützbarkeit historischer Datenbanken dienen, um zwei Ebenen geht, die häufig miteinander vermischt werden: die Überwindung technischer Unterschiede zwischen verschiedenen Softwareprodukten einer-, die präzise logische Beschreibung der getroffenen Codierentscheidungen andererseits. Zu diesem Punkt – der natürlich auch die im Gang befindliche Debatte über das derzeit optimale Maß an Quellentreue bei der Vorbereitung historischer Daten betrifft – ergreifen Jarausch und Schultz das Wort.

Abschließend sollen zwei Varianten der Verwendung von Datenbanken angeschnitten werden, die nur mehr mittelbar mit der Vorstellung von einem Instrument zu tun haben, das benützt wird um die auf eine eigene Forschungsaufgabe bezogenen Informationen zu verwalten. Braatz / Crane und Kropac stellen Möglichkeiten zur Verwendung von Datenbasen innerhalb der universitären Lehre vor. Rowney schließlich bespricht die Möglichkeiten, die

---

<sup>1</sup> Die Beiträge von Werner Braatz / Virginia Crane und Don Rowney wurden auf dem Welthistorikerkongreß in Stuttgart vorgelegt.

sich aus der Benützung der großen Informations- und Literaturlanken vom eigenen Schreibtisch aus für die Forschung ergeben.

Endlich hat der Herausgeber die angenehme Aufgabe, der Stiftung Volkswagenwerk nochmals für die Unterstützung der Veranstaltung, der die überwiegende Mehrzahl der hier abgedruckten Beiträge entstammt, zu danken.

Manfred Thaller

Göttingen, im Juni 1986



**Vorüberlegungen für einen internationalen Workshop  
über die Schaffung, Verbindung und Nutzung  
großer interdisziplinärer Quellenbanken  
in den historischen Wissenschaften.**

Der Gedanke, durch die Wiederverwendung einmal maschinenlesbar gemachter Daten im Rahmen anderer Projekte Ressourcen zu sparen, ist kaum jünger als die Heranziehung der Datenverarbeitung im geistes- und sozialwissenschaftlichen Bereich überhaupt. Eine große Anzahl von Projekten aller Disziplinen hat immer wieder darauf hingewiesen, daß ihre Bestrebungen auch und insbesondere unter dem Gesichtspunkt zu bewerten seien, daß durch sie exemplarische Daten für weit über die unmittelbaren Ziele hinausgehende Forschungsmöglichkeiten dem jeweiligen Fach zur Verfügung stünden. In den Sozialwissenschaften wurden aus derartigen Überlegungen eigene Einrichtungen der Infrastruktur der jeweiligen Fächer geschaffen, deren Ziel explizit die Archivierung und Aufbereitung von maschinenlesbaren Daten zur sekundären Analyse ist. In den Geisteswissenschaften (worunter wir hier und im Rest dieser Skizze die Literaturwissenschaften und die historischen Disziplinen verstehen) wurde diese Institutionalisierung trotz einzelner Ansätze kaum je erreicht, es gibt aber eine ganze Reihe von Einrichtungen, die auch hier die Archivierung und Bereitstellung maschinenlesbarer Daten wenn nicht als den zentralen, so doch als einen wesentlichen Bestandteil ihrer Funktionen ansehen.

Trotz dieser langen Tradition sekundäranalytischer Ansätze wird man die Bilanz tatsächlich durchgeführter sekundäranalytischer Arbeiten aber in wesentlichen Punkten als unbefriedigend empfinden müssen. Einzelne Forscher, die unter beträchtlichen Kosten und hohem persönlichen Einsatz spezialisierte Datenbanken bereitgestellt haben, klagten in aller Öffentlichkeit, daß die von ihnen geleisteten Vorarbeiten nur selten, wenn überhaupt genützt würden. Insbesondere im Bereich der historischen Wissenschaften überwiegt die Zahl zur Weiterverwendung angebotener Daten die dazu herangezogenen derzeit bei weitem.

Daß wir hier von einer in den mit Texten befaßten Disziplinen wesentlich unbefriedigenderen Bilanz sprechen, ist sicher nicht zufällig: einerseits ist selbst ein zusammenhängender Text von 100.000 Zeilen nur für einen unverhältnismäßig kleineren Teil der historischen oder literaturwissenschaftlichen Forschung von Interesse, als es eine in Form numerischer Codes vorliegende Datei gleicher Größe in den Sozialwissenschaften wäre. Andererseits variieren die Konventionen, nach denen Texte zur maschinellen Analyse auf-



bereitet werden, wesentlich stärker als die Eingabeformen statistisch orientierter Arbeiten, und hinzu kommt noch, daß die Analyse eines Textes von 100.000 Zeilen für die Ressourcen der meisten akademischen Rechenzentren bereits eine so große Belastung bedeutet, daß es kaum möglich wäre, mehrere andere Korpora gleicher Größe neben dem primär interessierenden nur zur Gewinnung von Vergleichswerten geringerer Signifikanz verarbeiten zu lassen. Demgegenüber bedeutet es bei der Geschwindigkeit derzeitiger Großrechner wenig, neben einer 2000er Stichprobe noch einige andere als Kontrolldateien zu verarbeiten, selbst wenn die dabei gewonnenen Ergebnisse gar nicht in die am Ende veröffentlichten Resultate eingehen sollten.

Im Rahmen der am Max-Planck-Institut für Geschichte laufenden Entwicklungsarbeiten an spezifisch auf die historische Forschung abgestimmten Softwaresystemen besteht der Eindruck, daß diese Situation sich vor allem für die historischen Wissenschaften im nächsten Jahrzehnt wegen einer ganzen Reihe bereits absehbarer technologischer Entwicklungen entscheidend ändern wird.

Grundsätzlich ist davon auszugehen, daß innerhalb weniger Jahre maschinenlesbare Quellen in ganz anderen Größenordnungen zur Verfügung stehen werden, als dies heute der Fall ist, und diese obendrein sehr viel effizienter bearbeitbar sein werden. Diese Entwicklung kann jedoch nur dann voll und fruchtbar genutzt werden, wenn rechtzeitig allgemeine infrastrukturelle Maßnahmen getroffen werden. Unter "maschinenlesbaren Quellen" verstehen wir alle maschinellen Darstellungsformen einer quellenorientierten Datenverarbeitung. Also sowohl Daten, die signifikante Textpartien zusammen mit einer großen Anzahl kodierter Angaben enthalten – wie sie in den Datenbanken der im weitesten Sinne demographischen Forschung vorliegen –, als auch solche, die fast nur Texte, diese aber in sehr stark strukturierten kleinen Einheiten umfassen – wie sie in den Datenbasen mikroanalytisch ausgerichteter Forschungsprojekte existieren –, und schließlich auch wenig kodierte Repräsentationen fortlaufender maschinenlesbarer Texte, wie sie bei computerunterstützten Editionstechniken entstehen. Unter einer Quellenbank verstehen wir eine Institution, die große Korpora derartiger Materialien unterschiedlichster Provenienz in integrierter Form so verwaltet, daß sie als Entitäten zwar faßbar bleiben, die gleichzeitige Bearbeitung in völlig verschiedenen Kontexten entstandener Quellen aber möglich wird.

In der derzeitigen Situation der Forschungsfinanzierung wäre es müßig, von der Gewinnung zusätzlicher Mittel für eine quantitative Ausweitung der vorhandenen Infrastruktur der historischen Forschung auszugehen. Es ist daher zu überlegen, wieweit eine verbesserte Zusammenarbeit bestehender infrastruktureller Einrichtungen und die gezielte Abstimmung einschlägiger

Bemühungen den sich anbahnenden Entwicklungen gerecht werden kann. Da die beschriebene Situation keineswegs auf ein Land beschränkt ist und die im folgenden angerissenen Probleme zum Teil überhaupt nur im internationalen Rahmen effizient lösbar sind, wären nationale Bemühungen von beschränktem Wert. Wie bereits betont scheinen uns die absehbaren Entwicklungen insbesondere für die datentechnisch arbeitenden Forschungsrichtungen innerhalb der Geschichtswissenschaften von Bedeutung: da die Lage in den Literaturwissenschaften jedoch sehr ähnlich und in deren historischen Zweigen nahezu mit der der historischen Disziplinen identisch ist, ist ihre Heranziehung für alle einschlägigen Überlegungen notwendig. Die längere Erfahrung und die wesentlich bessere einschlägige Infrastruktur der Sozialwissenschaften legt auch deren Einbeziehung nahe.

Als Bestandteil der Infrastruktur der Geschichtswissenschaft in der Bundesrepublik Deutschland schlug das Max-Planck-Institut für Geschichte daher einen internationalen Workshop von Vertretern einschlägiger infrastruktureller Einrichtungen sowie auf die Erzeugung großer Datenbanken ausgerichteter Forschungsprojekte vor. Das Institut erklärte sich gleichzeitig bereit, die Organisation einer derartigen Veranstaltung im Jahre 1985 zu übernehmen <sup>1</sup>.

Ziel dieser Veranstaltung sollte die Erarbeitung klarer Konzepte für mögliche Formen langfristiger und weitreichender Kooperation bei der Ausnutzung der durch die Datentechnik gebotenen Möglichkeiten sein, wobei diese Konzepte so konkret zu halten waren, daß ihre Implementation im Rahmen lokal laufender Entwicklungen möglich wird.

Wir möchten betonen, daß wir eine derartige Veranstaltung trotz der bewußt sehr weitgespannten Zielvorstellungen als für die *unmittelbare Arbeit* bedeutungsvoll ansahen. Um die einzurichtenden Arbeitsgespräche von vorneherein straff strukturieren zu können legten wir zunächst ein erstes Grundsatzpapier vor, das unsere Sicht der angesprochenen Themen darlegt und gleichzeitig als erstes Modell für die Strukturierung der vorgeschlagenen Arbeitsgespräche dienen sollte.

Wir gingen von folgenden Themenkreisen aus:

1. Welche Zielvorstellungen für eine optimale datentechnische Infrastruktur der historischen Disziplinen lassen sich aus den derzeit absehbaren technischen Entwicklungen ableiten?
2. Mit welchen organisatorischen Maßnahmen kann eine Annäherung an

---

<sup>1</sup> Die Kosten wurden zum größten Teil von der Stiftung Volkswagenwerk übernommen.

derartige Idealvorstellungen mit dem geringsten möglichen Aufwand erreicht werden?

3. Wieweit muß und kann sich eine Infrastruktur, die dem Fach maschinenlesbares Material in großem Maßstab zur Verfügung stellt, auch als Anregerin seiner Verwendung verstehen? Wieweit besteht dabei innerhalb einer europäisch verstandenen Infrastruktur die Möglichkeit zur Anregung comparatistischer Arbeiten?
4. Welche Möglichkeiten gibt es für ein internationales Ressource Sharing, also: (a) für die Übertragung spezieller Aufgaben der Informationserschließung innerhalb der Gesamtdisziplin an hardware- und/oder softwaremäßig dafür besonders gut ausgerüstete Institutionen? (b) für die Verrechnung derartiger Leistungen zwischen den nationalen Infrastrukturen?

### **Zielvorstellungen**

Der Einsatz der Datenverarbeitung in den Geisteswissenschaften zielte bisher in aller Regel auf die intensive Analyse einer Quelle. Dies unterschied vor allem in der Geschichtswissenschaft derartige Arbeiten vielleicht grundlegender von konventionellen, als der Einsatz der EDV an sich. Die Gründe dafür lagen hauptsächlich in der extrem arbeitsintensiven Vorbereitung, die für die Analyse einzelner Quellen notwendig ist, und in der Beschränkung der verfügbaren Ressourcen an den jeweiligen Rechenzentren, die der Größe selbst mit sehr einfachen Rechentechniken zu bearbeitender Korpora allein schon wegen der entstehenden I/O Zeiten eine Grenze setzten.

In herkömmlichen Forschungsansätzen konzentriert man sich zwar üblicherweise ebenfalls auf ein überschaubares Korpus: gleichzeitig wird aber eine sehr viel größere Anzahl von Quellen für einzelne zusätzliche Argumente herangezogen. Dadurch erscheinen selbst vom heutigen datentechnischen Stand sehr große Datenbanken als relativ eng umgrenzte Quellengruppe, wenn sie mit der Ausdehnung der traditionellen Arbeiten zugrundegelegten Quellen verglichen werden. Als signifikantes Beispiel dafür mag die Reaktion stehen, die die Bemerkung des Verfassers, integrierte Datenbanken, die den gesamten Bestand der Monumenta Germaniae Historica (nach einer groben Schätzung bereits als unstrukturierte Eingabedaten etwa 0.5 Gigabyte) umfaßten, wären heute mit großem, aber absehbaren Aufwand realisierbar, in einem Kreis überwiegend ohne EDV arbeitender Historiker hervorrief: zweifellos, so wurde bemerkt, handle es sich dabei um eine sehr zentrale Quellengruppe; im Falle des Diskussionsredners müßte eine Datenbank, um alle für seine Arbeit relevanten Quellen aufzunehmen, aber etwa die 300fache Menge an Material enthalten.

Als illustratives Beispiel mag dies zeigen, daß man bei Designüberlegungen für eine Infrastruktur, die innerhalb der historischen Disziplinen von zahlreichen Forschungsprojekten mit wirklichem Gewinn genützt werden kann, um Größenordnungen anders denken muß als bisher. Kein Forscher kann freilich in Anspruch nehmen, eine derartige Materialmenge mit der gleichen Intensität bei der Auswertung des Informationsgehaltes ihrer elementaren Bestandteile zu bearbeiten, wie dies etwa bei der statistischen Auswertung dazu geeigneter Quellen geschieht. In aller Regel wird aus derartigen Korpora ein sehr viel kleinerer Teil des Materials zur eigentlichen Bearbeitung herangezogen, wobei auch bei konventioneller Arbeitsweise die Phasen der Quellensichtung und der Analyse des dadurch erhaltenen Materials unterscheidbar sind.

Will man die durch derartige Bemerkungen gebotene Herausforderung an eine datentechnische Infrastruktur in den historischen Disziplinen annehmen, ergibt sich als erste Zielvorstellung ein Mechanismus, der eine nach heutigen Vorstellungen extrem große Datenmenge – jedenfalls mehrere Gigabyte – verwaltet, aus der dann von Fall zu Fall mittels geeigneter analytisch “flacher” Zugriffsmechanismen relevante Teile herausgezogen und als “Metaquelle” für die Bearbeitung mittels sehr viel tiefergehender analytischer Routinen zusammengestellt werden.

Jede Umsetzung dieser Zielvorstellung wird sich fünf grundlegenden Problemen gegenüber sehen:

- a. dem primären Zugang zu maschinenlesbaren Quellen in großer Zahl,
- b. ihrer Sicherstellung in einer Weise, die den Zugriff auf Teile davon im Bedarfsfall mit einem Minimum an menschlicher Intervention jederzeit kurzfristig erlaubt, ohne die permanente Residenz des Gesamtmaterials auf Random Access Speichern zu verlangen,
- c. ihrer Sicherstellung in neutraler, also dem Erscheinungsbild der Quelle ohne Zwischenschaltung konzeptueller Vorentscheidungen maximal angenäherter Form,
- d. der Schaffung eines Selektionsmechanismus, der in der Lage ist, unterschiedliche Teile des Gesamtmaterials mit unterschiedlicher Eindringtiefe zu bearbeiten, um rechenaufwendiges Suchen in sicherlich irrelevanten Quellen zu vermeiden, und
- e. der Notwendigkeit, aus dem Ausgangsmaterial selektierte Teilmengen in einer großen Anzahl extrem unterschiedlicher Darstellungsformen weiterverarbeiten zu können.

Die damit angeschnittenen Probleme sind, wenn man große Quellenmengen betrachtet, sicherlich heute noch nicht endgültig und definitiv lösbar. Un-

seres Erachtens gibt es jedoch in jedem der angeschnittenen Problemkreise Entwicklungen, die uns jetzt schon zumindest in die Lage versetzen, mittel- bis langfristig gangbare Lösungswege aufzuzeigen.

Unsere Zielvorstellungen für eine integrierte Quellenbank großen Ausmaßes lauten:

Um die Bereitstellung maschinenlesbarer Datensätze für die tägliche Arbeit in der historisch orientierten Forschung wirklich wirksam werden zu lassen, muß von der Verfügbarkeit unverhältnismäßig größerer Datenmengen als bisher ausgegangen werden.

Aufgabe einer als infrastruktureller Einrichtung verstandenen Quellenbank ist zunächst das Sammeln aller einschlägigen Materialien. Solche fallen an als Überbleibsel abgeschlossener Forschungsprojekte im klassischen sekundäranalytischen Sinn, in der Form von maschinenlesbaren Drucksätzen von Quelleneditionen und durch gezielte Umwandlung großer Korpora mittels mechanischer Lesegeräte <sup>2</sup>.

---

<sup>2</sup> Hier und im folgenden sollen unsere technischen Überlegungen als Anmerkungen folgen, um den programmatischen Charakter der engeren Zielvorstellungen nicht zu stören.

Maschinenlesbares Material, das für die historische Forschung von Interesse ist, fällt heute und in der näheren Zukunft aus drei Quellen an:

1) Bei Projekten, die sich bereits heute der Datenverarbeitung im eingangs skizzierten Bestreben bedienen, ein vergleichsweise kleines Quellenmaterial intensiv zu analysieren.

2) Als Resultat der Druckaufbereitung. Der rapide Vormarsch der Elektronik im Druckwesen liefert bereits heute maschinenlesbare Äquivalente vieler Quellenpublikationen. Da anzunehmen ist, daß sich diese Tendenz bereits in der näheren Zukunft noch erheblich verstärkt, kann man davon ausgehen, daß in absehbarer Zeit jede überhaupt neu edierte Quelle maschinenlesbar vorliegt. Dies gilt noch wesentlich stärker für eine primär statistisch/quantitativ verstandene Forschung; freilich bilden die sich aus dem Datenschutz ergebenden Fragen der Zugänglichkeit primär statistischer Materialien ein besonderes Problem, das aber eher in den organisatorischen als technischen Bereich zu verweisen ist.

3) Schließlich auf Grund neuer Technologien billig zugänglich werdende Materialien. Ohne endgültig entscheiden zu können, ob die Geräte bei ihrem derzeitigen technischen Stand bereits in der Lage sind, ältere, im Bleisatz erstellte Texte in jedem Fall schnell aufzubereiten, kann man doch davon ausgehen, daß die Verfügbarkeit von Kurzweil Data Entry Machines oder ähnlichen Geräten innerhalb kurzer Zeit

Erstes Ziel einer derartigen Einrichtung ist die Bereitstellung der gesamten Quellenbank in einer abgestuften Speicherungsform, die das Gesamtmaterial in mehrere Ebenen auf die zur Verfügung stehenden Speichermedien verteilt. Data Dictionaries und andere Behelfe, die zur ersten Beurteilung von einzelnen Quellen/Dateien benötigt werden, sind dabei sofort zugänglich; andere benötigte Komponenten des Gesamtbestandes, insbesondere der volle Quellentext, werden bei Bedarf vom System automatisch von geeigneten "unbeschränkten" Hintergrundmedien geladen <sup>3</sup>.

---

die Kosten für die Lesbarmachung sehr vieler Quellen radikal senken und auch den Gedanken an die gezielte Aufbereitung großer Korpora, ohne sie sofort mit einer eng verstandenen Analyseabsicht zu verbinden, möglich machen wird.

In allen drei Fällen entstehen freilich Daten, die von Projekten, die an Teilen ihres Informationsgehalts interessiert sind, nicht ohne vorbereitende Arbeiten verwendet werden können. Teil der zu skizzierenden Infrastruktur muß also Software sein, die in der Lage ist: (erstens) von einzelnen Projekten erstellte, quellengetreue Daten rasch und mit einem Minimum an menschlichem Eingreifen in eine neutrale, untereinander austauschbare Form umzusetzen, (zweitens) die zur Speisung von Satzmaschinen erzeugten Magnetbänder und andere Datenträger rasch so umzuformen, daß die satzrelevanten Flags entfernt werden und der verbleibende Text ebenfalls auf die bereits erwähnte neutrale Form abgebildet wird, und (drittens) entsprechende Leistungen für mit einem Minimum an Aufbereitung von mechanischen Lesegeräten produzierte maschinenlesbare Texte zu liefern.

<sup>3</sup> Für die Speicherung mit kurzfristigen Zugriffszeiten wird in Eröffnungsreden auf einschlägigen Kongressen gerne global auf die rapide sinkenden Preise sekundärer Speichermedien verwiesen. Wir glauben, daß wir es uns mit einem derartigen Verweis, wenn wir in absehbarer Zukunft mit der Schaffung einer Infrastruktur beginnen wollen, die Datenmengen der angegebenen Größenordnung bewältigen soll, viel zu leicht machen würden. Unseres Erachtens wird in der planbaren Zukunft selbst eine noch so drastische Reduktion der Preise für Speichermedien nicht in der Lage sein, mit der vorhersehbaren Explosion für die Geisteswissenschaften relevanter maschinenlesbarer Daten Schritt zu halten. Wir glauben, daß sich aber auch hier Lösungen abzeichnen, die sehr wohl mit auf der Reduktion der Speicherpreise beruhen.

Wir gehen davon aus, daß die entscheidende Frage ist, wieweit es gelingt, Daten in Form von hierarchisch auf verschiedenste Speichermedien abgelegten diskreten Einheiten, die untereinander durch ein internes Referenzsystem verbunden sind, so abzulegen, daß der Zugriff auf sie keine Kenntnisse des Benützers über die Eigenheiten des Betriebssystems des Rechners voraussetzt. Im Prinzip gehen wir dabei davon aus, daß heute in sich abgeschlossene Quellen (die Transkription ei-

Die Sicherstellung der Quellentreue der verspeicherten Materialien liegt im Ermessen der die einzelnen Quellenbanken unterstützenden Institutionen und bildet Teil des Designs einzelner Entwicklungen: in jedem Fall ist jedoch eine Möglichkeit vorzusehen, die zur Verfügung stehenden Materialien auf ein internationales Austauschformat abzubilden <sup>4</sup>.

---

nes bestimmten Bestandes administrativer Aufzeichnungen, ein maschinenlesbar gemachter fortlaufender Text) auch vom Rechner noch als sinnvolle Entitäten erkannt werden. Wir sehen vor, daß auf dem verfügbaren Random Access Speicher nur relativ kleine Teile komplexer Data Dictionaries resident sind, die alle notwendigen Verweise enthalten, um benötigte Entitäten ohne bewußtes Eingreifen des Benützers von im Prinzip unbegrenzten Datenträgern (wie etwa einer Magnetbandbibliothek) zur Bearbeitung laden lassen zu können.

Wir möchten betonen, daß diese Lösung bereits heute arbeitsfähig ist (in der Tat im Max-Planck-Institut für Geschichte in relativ einfacher Form eben realisiert wird), aber mit der Entwicklung neuer Speichermedien - etwa der Verfügbarkeit großer fix montierter Laserdisks zum teilweisen Ersatz heutiger Plattensysteme oder der Verfügbarkeit wechselbarer Laserdisks zum Ersatz von Bandeinheiten - linear fortgeschrieben werden kann, da sie bloß von der Unterteilung des gesamten verfügbaren Speichermediums in mehrere Klassen mit unterschiedlichen Zugriffscharakteristika ausgeht. Daß Data Dictionaries und File Directories selbst hierarchische Gebilde formen können, ist angesichts der Entwicklung der letzten Jahre auf dem Gebiet der Betriebssysteme ja wohl selbstverständlich.

<sup>4</sup> Die Sicherstellung des Materials in einer neutralen Form, die dem Erscheinungsbild der Quelle maximal angenähert ist, ohne durch konzeptuelle Vorentscheidungen beeinflusst zu sein, ist vielleicht die schwierigste Aufgabe überhaupt.

Nicht so sehr technisch gesehen. Hier liegen zwar ebenfalls einige, keineswegs triviale Probleme vor, schließlich soll ja gewährleistet werden, daß das Material zwar das Erscheinungsbild der Quelle wiedergibt, gleichzeitig aber durch die noch zu diskutierenden Selektionsmechanismen effizient bearbeitet und für Auswertungen leicht umgeformt werden kann. Man wird also kaum umhinkönnen, gleichzeitig mit dem einfachen Quelleninhalt Strukturierungshilfen zu verspeichern, die das Material nicht in konzeptuellem, sehr wohl aber in technischem Sinne vorgliedern und von vorneherein auch eine Reihe von Filtern vorsehen müssen, mit deren Hilfe sich gegebenenfalls der verspeicherte Quelleninhalt auf die relevanten Teile einschränken läßt. Ein Beispiel für letzteres Problem ergibt sich etwa bei der Bearbeitung erzählender historischer Quellen: soll derartiges Material für die sprachwissenschaftlich orientierte Analyse in eine entsprechende "Metaquelle" im oben definierten Sinn umgewandelt werden, wird man Wert darauf legen, vorhandene diakritische Zeichen als Indikatoren für Lautlängen und ähnliches voll zu bewahren.

Primäres Ziel einer Quellenbank im hier beschriebenen Sinn ist die Anlage eines möglichst großen Korpus für die historischen/literaturwissenschaftlichen Disziplinen insgesamt, vor allem die Verhinderung der Zerstörung bereits maschinenlesbarer Materialien, nicht so sehr ihre analytische Durchdringung. Um diese zu erleichtern, ist jedoch dafür Sorge zu tragen, daß bestimmte Klassen einmal durchgeführter rechenintensiver Auswertungen innerhalb des Gesamtsystems bekannt bleiben und bei ähnlichen Anfragen einem Benutzer ohne menschliches Eingreifen als "billigere" Einstiege in sein Forschungsproblem angeboten werden. Durch die Kumulation derartiger Resultate und zusätzliche Bemühungen um die Aufnahme globaler Angaben über einzelne Quellen besteht die Möglichkeit, das Gesamtkorpus in mehre-

---

Soll die Auswertung desselben Materials nach sozialhistorischen Gesichtspunkten erfolgen (etwa durch die Selektion aller mit Personennamen verbundenen Textteile und die Verbindung dieser Textteile mit andersgearteten, primär statistisch relevanten Quellen), wird man die nachfolgenden Arbeiten stark vereinfachen, wenn man derartige Auszeichnungen so früh wie möglich entfernt.

Dennoch glauben wir, daß diese technische Seite der Probleme durchaus in den Griff bekommen werden kann; wesentlich gravierender scheint uns die organisatorische Seite. Rein technisch gesehen wäre die verbindliche Vereinbarung eines hochkomplexen internationalen Standards für die langfristige Speicherung großer textorientierter Datenmengen unter besonderer Berücksichtigung der Eigenarten historischen Quellenmaterials die vernünftigste Lösung. In der Praxis müssen wir nach allen Erfahrungen aber davon ausgehen, daß jene Institutionen, die einschlägige Arbeiten schon geleistet haben, nur in den seltensten Fällen bereit sind, auf die Beibehaltung der von ihnen entwickelten Standards zu verzichten; zum Teil allein schon deshalb, weil die Finanzierung in vielen Ländern wesentlich von der Vorstellung der Finanzierungsträger bestimmt ist, im jeweils eigenen Land völlig neue Maßstäbe für die internationale Entwicklung zu setzen.

Man wird also auf die technisch optimale Lösung verzichten müssen.

Machbar und wünschenswert ist dagegen die Vereinbarung eines maschinenunabhängigen Austauschformates für in große Quellenbanken zu verspeicherndes Material, etwa im Sinne der Metafiles der GKS Normungsvorschläge. Unter der Voraussetzung, daß die am Datenaustausch interessierten Institutionen in der Lage sind, einerseits die von ihnen intern in einer Vielzahl von Basisformaten gespeicherten Rohdaten (gegebenenfalls unter Verlust lokaler Besonderheiten) zum Austausch in das Normformat zu übertragen und andererseits in diesem Format empfangene Daten (gegebenenfalls unter Auslassung örtlich als nicht relevant angesehener Angaben) in das jeweils intern gültige zu übertragen, ist der größte Teil der anfallenden Probleme bereits gelöst.



ren, durch ihre analytische Tiefe und den dadurch bedingten Rechenaufwand unterschiedenen, Stufen nach in ein konkretes Projekt einzubeziehenden Teilen zu durchsuchen<sup>5</sup>.

---

<sup>5</sup> In der voraussehbaren Zukunft ist wohl kaum davon auszugehen, daß die vollständige Durchsuchung mehrerer Gigabyte nach bestimmen Suchbegriffen eine realistische Lösung für ein häufig einzusetzendes Arbeitsinstrument ist. Unseres Erachtens liegen Lösungsmöglichkeiten in einer Verfeinerung der bereits diskutierten Datenorganisation.

Zunächst denkt man natürlich an partielle Inversionen der Teilbestände in der gesamten Quellenbank, die wir uns ja als eine aus zahlreichen kleineren Datenbasen zusammengesetzte übergeordnete Datenbank vorstellen können. Für viele häufig verfolgte Forschungsinteressen wäre dies auch sicher eine Lösung. Allgemein wird man bei einer Quellenbank, die in der Lage sein sollte, gleichmäßig eine große Anzahl von Teildisziplinen zu bedienen, aber kaum in der Lage sein, alle relevanten Aspekte in partielle Inversionen einzugliedern. Zudem bleibt – auch wenn wir uns vorerst nur um Zielvorstellungen bemühen und die Frage ihrer Realisierbarkeit zunächst verschieben – jetzt schon festzuhalten, daß eine derartige Konzeption von einer intensiven Bearbeitung jeder einzelnen Teilquelle durch die die gesamte Quellenbank bereithaltende Institution ausgeht, was mit vorhandenen Ressourcen kaum realisierbar wäre. Unseres Erachtens müßte die Lösung in einer Doppelstrategie gesucht werden:

Einerseits in einer Erweiterung der Funktion der in das Gesamtsystem zu integrierenden Data Dictionaries, die nicht nur innerhalb der einzelnen verwalteten Quellen die elementaren Einheiten (im Sinne von "Variablen") beschreiben, sondern auch so viel wie möglich an Angaben über die Gesamtquelle beinhalten – Entstehungszeit, behandelter Zeitraum, sprachliche Charakteristika, eine kurze Quellenbeschreibung, eine Charakterisierung ihrer Bedeutung für die einzelnen Subdisziplinen etc. etc. Dementsprechend würde der Zugang zu den einzelnen Quellen stufenweise erfolgen – aber mit den erweiterten Möglichkeiten der Datenverarbeitung auf *allen* Ebenen und unter Integration von Suchbehelf und Quellenbestand.

Andererseits ist neben dieses Prinzip einer hierarchischen Datendokumentation noch das der Additivität der Anstrengungen einzelner Benützer zu setzen. Wir gehen davon aus, daß in der Regel einmal in den Basisbestand aufgenommene einzelne Quellen kaum mehr verändert werden, daß sie bloß Teile an von einzelnen Benützern sukzessive aufgebaute Metaquellen abgeben, die dann ihrerseits, als eigentliche Datenbasen der einzelnen Forscher, häufigen und intensiven Veränderungen unterliegen. Dies bedeutet, daß alle einmal für den Basisbestand durchgeführten Berechnungen valide bleiben. Dementsprechend schiene es sinnvoll, zumindest für bestimmte Klassen von Bearbeitungen (wie etwa die Erstellung partieller Inversionen

Neben diesen grundlegenden Fähigkeiten bietet jede Quellenbank die Möglichkeit, ausgewählte Teile des Gesamtkorpus in eine "Metaquelle" in einer Vielzahl von Formen auszugeben, die der analytischen Weiterverarbeitung durch die jeweils besonders unterstützten Ansätze entgegenkommen<sup>6</sup>.

Soweit zu unseren Zielvorstellungen für die *Struktur* einer idealen, d.h. die absehbaren technischen Möglichkeiten voll nützenden "großen" Quellenbank. Wir haben bisher bewußt noch nichts über die unserer Ansicht nach anzustrebende *Organisation des Zugangs* zu einer derartig strukturierten Einrichtung gesagt, da wir glauben, daß hier eine ganze Reihe verschiedener Formen gleichberechtigt nebeneinander vorzusehen sein wird.

Prinzipiell gehen wir davon aus, daß eine Quellenbank, die sich als Beitrag zur Infrastruktur ihres Faches versteht, auf einem Großrechner eines akademischen Rechenzentrums installiert wird. Ohne hier auf die nationalen Unterschiede in der Finanzierung datentechnischer Dienstleistungen für die geisteswissenschaftlichen Disziplinen eingehen zu können oder zu wollen, setzen wir voraus, daß es das *Ziel* jeder derartigen Einrichtung sein muß, ihre Dienste innerhalb mindestens des nationalen Rahmens auf einer ähnlichen Ebene, wie dies wissenschaftliche Bibliotheken tun, zur Verfügung zu stellen, also im wesentlichen kostenlos.

Dabei ist die Datenhaltung so zu organisieren, daß ein benützender Forscher die oberen Schichten von Selektionsvorgängen (also die Durchsuchung nicht von Quellen, sondern von Quellenbeschreibungen) jederzeit ad hoc vornehmen kann, während jede im System verspeicherte Quelle innerhalb von 6 Stunden zur intensiveren Bearbeitung zur Verfügung steht.

Ein Verbund derartiger Quellenbanken ist vorzusehen, der im nächsten Hauptpunkt noch zu diskutieren sein wird.

---

einzelner Quellen innerhalb des Gesamtbestandes) eigene Dictionaries anzulegen. Dies versetzt das Gesamtsystem in die Lage: (a) derartige Berechnungen erst dann durchzuführen, wenn sie wirklich notwendig werden, (b) ihre Ergebnisse danach für alle späteren Benützer sicher zu stellen und (c), in Verbindung mit einem etwas besser ausgebauten Logging System (also einer "Tagebuchführung" des Gesamtsystems), Benützer von ihren Interessen wechselseitig in Kenntnis zu setzen.

<sup>6</sup> Die Frage, welche Möglichkeiten zur Erstellung von analyseorientierten Dateien aus der quellenorientierten Gesamtdatenbasis bestehen, wird für die praktische Verwertbarkeit jeder konkreten Quellenbank von ganz zentraler Bedeutung sein – gleichzeitig bildet dies aber ein Problem, das erst für die Designüberlegungen konkret zu realisierender einzelner Quellenbanken von Bedeutung ist, dagegen kaum entscheidend für die Überlegung, ob Quellenbanken als solche realisiert werden können und was die Zielvorstellungen für sie sein müßten.

Die zentral unterhaltenen Quellenbanken stellen interessierten Institutionen Teile des bereitgehaltenen Materials zur Verfügung, setzen dabei aber auch voraus, daß bei diesen Institutionen aufbereitete Daten und als speicherungswürdig klassifizierte Analysen ihnen wieder zugänglich gemacht werden.

Dies bedingt grundsätzlich eine Vernetzung der zentralen Quellenbank mit den Rechnern der Teile des Gesamtkorpus bearbeitenden Institutionen. Wir halten die energische Erinnerung daran für nötig, daß wir hier keine Festtagsreden halten, sondern über Konzepte sprechen, mit deren praktischer Verwirklichung sofort begonnen werden kann. Wir halten Spekulationen über die Zeiten, da man in der Lage sein werde, Quellenbanken, wie wir sie hier beschreiben, via Satellit in allen Teilen der Welt on line zu bearbeiten, im derzeitigen Stand der Entwicklung für weitgehend unsinnig.

Wenn wir von einer "Vernetzung von Datenbanken" sprechen, halten wir uns vor Augen, daß:

- insbesondere in den Anfangstagen einer entstehenden Quellenbank die Zahl der Benützer sehr klein und der Abstand zwischen einzelnen Anfragen sehr groß sein wird. Die Anschaffung von Hardware, die nur gelegentlich im Abstand von Tagen benützt wird, halten wir für nutzlos und gefährlich, da sie nur den prinzipiellen Gegnern der Anwendung formaler Verfahren in den historischen Disziplinen Argumente liefert.
- das Konzept einer *Quellenbank* – zum Unterschied von einer *Datenbank* – die Übertragung ganzer Korpora von Quellen vorsieht und daher aus Kapazitätsgründen die Verwendung heute und in absehbarer Zeit zur Verfügung stehender kurzfristig anwählbarer Übertragungsleitungen der Postverwaltungen ausschließt.
- bei der Arbeit mit dem hier behandelten Material voluminöse Resultate zu erwarten sind. Eine "rough and ready concordance" z.B. erscheint uns am Bildschirm nicht eben als sinnvoll darstell- und beurteilbar.

Für die Kommunikation zwischen einer zentralen Quellenbank und einzelnen Benützern, die daraus entnommene Metaquellen mittels lokaler Datenbanken verwalten, schlagen wir folgende Ebenen vor:

- Für die Weitergabe einzelner Korpora und die Rückgabe der Ergebnisse bestimmter rechenaufwendiger Analysen den Versand der in Frage kommenden Daten auf Magnetbändern oder anderen externen Datenträgern<sup>7</sup>. Dabei ist von einem innerhalb des Netzwerkes der jeweiligen Quellenbank verbindlichen Austauschformat auszugehen, das von

---

<sup>7</sup> Unseres Erachtens liegt das Problem bei der bisherigen Organisation von Archiven maschinenlesbarer Daten nicht darin, daß der Versand von Magnetbändern ein so extrem ineffizienter Weg wäre. Es liegt vielmehr darin, daß vor allem die

entsprechenden sekundären Datenbanksystemen unmittelbar verarbeitet werden kann und eine extensive Datenbeschreibung in standardisierter Form gleich mitläßt. Diese Angaben sind jedoch so zu halten, daß sie, wenn ein derartiges System *nicht* zur Verfügung steht, gegebenenfalls mit einem Editor einfach vom eigentlichen Text getrennt werden können. Auf dieser Ebene ist dann auch an die Möglichkeit zu denken, nicht nur Quellen, sondern auch Teile der Quellenbeschreibungen auszulagern und in sekundären Datenbanken zugänglich zu machen.

- Eine unmittelbare Vernetzung scheint uns derzeit nicht so sehr zwischen einer einzelnen Quellenbank und von ihr abhängigen Datenbanken von Nutzen, sehr wohl aber als eine Möglichkeit, die arbeitsintensive Verwaltung der Quellenbeschreibungen auf mehrere Institutionen aufzuteilen. Die Realisierung dieses Konzeptes kann etwa bedeuten, daß zwei Quellenbanken einerseits auf die literarischen Quellen einer bestimmten Epoche, andererseits auf die historischen spezialisiert sind. Die eigentlichen maschinenlesbaren Quellen liegen dann in den Hintergrundmagnetbandbibliotheken *beider* Institutionen; jede Institution verwaltet aber nur die Quellenbeschreibungen und Data Dictionaries ihres Fachgebietes und hat die Möglichkeit, via Netz (eventuell ohne daß der Benützer dies merkt) die an der anderen Institution verwalteten zu konsultieren, wonach gegebenenfalls die benötigten Quellen aus der eigenen Hintergrundbibliothek geladen werden können.
- Wir halten für eine weitere Möglichkeit der Vernetzung in Verfolgung des eben Gesagten die Zeit für Detailüberlegungen noch nicht für reif, wollen aber doch auf sie verweisen. Grundsätzlich muß es unserer nach in absehbarer Zeit möglich sein, ein System von Quellenbeschreibungen und Data Dictionaries in erheblichen Teilen in ein Netz voneinander völlig

---

kleineren derartigen Archive Anfragen sehr selten bekommen und es daher u.U. Monate, wenn nicht Jahre dauert, bis eine angeforderte Quelle auf einem Magnetband lokalisiert und aus einer mittlerweile obsoleten Aufzeichnungsart in eine der jetzigen Anlage verständliche gebracht wird, bevor die Erstellung des andernorts lesbaren Bandes ins Auge gefaßt werden kann. Und es liegt ferner darin, daß in manchen Fällen die bei der Datenerfassung verwendeten Kodiersysteme selbst der archivierenden Institution nicht mehr bekannt sind, ganz zu schweigen von sonstigen Informationen über das Material. Beide Probleme - denen zumindest das Max-Planck-Institut für Geschichte in der Sammeltätigkeit für ein einschlägiges Projekt immer wieder begegnete - können aber durch noch so intensive Bemühungen um die neuesten technischen Errungenschaften nicht gelöst werden, sondern liegen ganz auf der Ebene der Organisation.

unabhängiger Micros zu verlagern, während die eigentliche Quellenbank auf einer weitgehend als höchst intelligente Memory Bank verstandenen Großrechenanlage gehalten wird.

## **Derzeitige organisatorische Möglichkeiten**

Viel im vorangegangenen Abschnitt Gesagtes mag heute nahezu utopisch erscheinen; wir sind jedoch der Ansicht, daß organisatorische Initiativen auf dem Weg zu diesem Ziele jetzt schon möglich und auch sehr wesentlich sind, wenn vermieden werden soll, daß die fachspezifischen Belange der Geisteswissenschaften bei der Entwicklung der an Universitäten zur Verfügung stehenden informationstechnischen Infrastruktur weiter ins Hintertreffen geraten. Folgende Maßnahmen sind u.E. sofort – und mit einer Ausnahme ohne Ausweitung der bestehenden Infrastruktur – realisierbar und sinnvoll.

- a. Die Intensivierung der Sammlung einschlägig erstellter Datenbasen. Es hat zwar bereits zahlreiche Bemühungen um die zentrale Sicherstellung für eine Teildisziplin aufbereiteten Quellenmaterials gegeben, die Erfahrungen im Max-Planck-Institut für Geschichte – sowohl beim eigenen Bemühen um Daten, als auch im Gespräch mit Kollegen, die ihre Daten gerne zentral archivieren lassen würden – zeigen aber, daß diese Dienste noch nicht hinreichend bekannt sind.
- b. Die Sicherstellung der elektronischen Repräsentationen aller mit Hilfe des Computers gesetzter Werke. Hier ist u.E. zwischen zwei Problemen zu unterscheiden:
  - einerseits hat es Fälle gegeben, wo die Satzunterlagen von Editionen, bei denen eine Zweitauflage unwahrscheinlich war, vom Verlag vernichtet wurden. Dem müßte durch eine rechtzeitige Sammlung derartiger Materialien vorgebeugt werden.
  - andererseits gibt es Verlage, die unter Berufung auf das Copyright Ansprüche auf die elektronische Repräsentation eines Werkes sogar dann erheben, wenn das die Satzmaschinen steuernde Magnetband vom Verfasser an einem akademischen Rechenzentrum selbst erstellt wurde.

Insbesondere die letzte Situation sollte geklärt werden, bevor sie sich mit unabsehbaren Folgen zum Gewohnheitsrecht verfestigt. Nach Ansicht des Verfassers sollte grundsätzlich festgeschrieben werden, daß die elektronischen Satzunterlagen von Werken mit Editionscharakter einschlägigen infrastrukturellen Einrichtungen zur Verfügung gestellt werden müssen. Einen Nachdruck von einer derartigen Darstellung der Quelle durch Dritte zu verhindern, kann gesetzlich nicht schwer sein,

sodaß die legitimen Interessen des Verlagswesens keinesfalls tangiert werden: das ebenso legitime Interesse der Wissenschaft, eine einmal – letztlich auf Kosten der öffentlichen Hand – finanzierte Edition in bestmöglicher Form benutzen zu können, und das heißt eben auch unter Einsatz der elektronischen Datenverarbeitung, darf von allem Anfang an nicht zu einer Umwegsubventionierung mißbraucht werden.

- c. Die eine Ausnahme von der Forderung, nicht über zusätzliche infrastrukturelle Mittel zu sprechen, sondern über die bestmögliche Ausnützung bestehender, bildet der Vorschlag, im Rahmen einschlägiger Institutionen die Frage zu besprechen, wie weit es möglich ist, sich gezielt um die langfristige Aufbereitung großer Quellenkorpora unter Einsatz mechanischer Lesegeräte zu bemühen.

Unserer Ansicht nach haben Quellenbanken, wie wir sie hier besprechen, wesentlich langfristige Bedeutung. Es wäre kindisch – und hat unserer Ansicht nach vergleichbaren Vorschlägen in der Vergangenheit nur geschadet –, so zu tun, als würde die bloße Gründung einer Quellenbank innerhalb der unmittelbar darauf folgenden Jahre zu einer explosionsartigen Vermehrung sie benützender Forschungsprojekte führen. Deshalb sollte der Einsatz eines derartigen Gerätes zur systematischen Aufbereitung großer Korpora, wobei in erster Linie wohl die klassischen Editionsreihen in Frage kämen, nur unter folgenden Rahmenbedingungen angestrebt werden:

- ein derartiges Gerät sollte grundsätzlich neben der langfristigen Umwandlung eines Korpus für ad hoc anfallende Wünsche der Disziplin zur Verfügung stehen,
  - dabei wäre für die langfristige Komponente aber von vornherein ein bestimmtes Kontingent der Gesamteinsatzzeit zu sichern.
  - die Umwandlung sollte nicht nur auf die Speicherung und Bereithaltung der Quellen abgestimmt werden, sondern gleichzeitig auch Grundlage der unmittelbaren Schaffung von Arbeitsbehelfen für die in Frage kommenden Korpora sein.
  - im Idealfall sollte von Anfang an ein inhaltliches Projekt, das das aufbereitete Material sofort verwendet, mitlaufen.
- d. Die sich konstituierenden Quellenbanken sollten grundsätzlich im freien Austausch miteinander stehen und im Prinzip alle den gleichen Bestand an maschinenlesbaren Quellen bereithalten. Ganz abgesehen vom eher trivialen Vorteil einer dadurch gegebenen absoluten Datensicherheit scheint uns dies einerseits wünschenswert, da dadurch die Möglichkeit zum Leistungsvergleich konkurrierender Designs gegeben ist, andererseits weil dadurch ein erster wichtiger Grundstock für eine überregionale

Vernetzung im bereits besprochenen Sinn gegeben wäre. Als Voraussetzung für die Realisierung dieses Prinzips ist davon auszugehen, daß der Datenaustausch zwischen derartigen Institutionen grundsätzlich kostenfrei erfolgt.

- e. Vor allem der letzte Satz, aber auch generell die Punkte a bis d erfordern die Regelung der Fragen des Datenaustausches durch allgemein als verbindlich anerkannte Richtlinien. Die beiden leitenden Gedanken müssen sein, daß (erstens) im Bereich der Forschung einmal erbrachte Leistungen der gesamten Disziplin ohne weitere Kosten erhalten bleiben müssen, daß aber (zweitens) andere daraus keinen monetären Gewinn ziehen dürfen. Es ist uns klar, daß dies im wesentlichen bereits jetzt der herrschenden Praxis entspricht. Andererseits ist aber immer wieder zu beobachten, daß bei der Anfrage nach Daten, die in allen einschlägigen Publikationsmedien als "auf Anforderung verfügbar" bezeichnet werden, einerseits recht zeitraubende Prozeduren notwendig werden, um die Genehmigung der ihre Aufbereitung ursprünglich finanzierenden Körperschaft einzuholen, andererseits in einzelnen Fällen durchaus auch versucht wird, einmal aufgelaufene Projektkosten später durch einen "Verkauf" der Daten zumindest zum kleinen Teil wiederzuerlangen.

Wir halten dies auch im Interesse aller datentechnische Methoden propagierenden Forscher in den historischen Wissenschaften für gefährlich, weil die einschlägigen Methodologien sich erst dann wirklich verbreiten werden, wenn es auch und gerade für studentische – d.h. also extrem ressourcenarme – Projekte möglich wird, sich ihrer zu bedienen.

Deshalb schlägt das Max-Planck-Institut für Geschichte zur Regelung der angesprochenen Probleme im Zusammenhang mit den Punkten (a) bis (d) eine Übereinkunft der entsprechenden infrastrukturellen Einrichtungen des europäischen Raumes an Hand folgender Punkte vor:

1. Maschinenlesbare Darstellungen historischer Quellen, die sich im Besitz einer Quellenbank befinden, stehen der Disziplin grundsätzlich kostenlos zur Verfügung.
2. Jede Weitergabe von maschinenlesbaren Quellen ist an eine schriftliche Einverständniserklärung gebunden:
  - sie ausschließlich zu wissenschaftlichen Zwecken zu verwenden,
  - keine Publikationen daraus zu produzieren, die Partien der maschinenlesbaren Repräsentation erneut in maschinelle Satzanweisungen umsetzen (um z.B. Anthologien unter Umgehung des Copyrights zu erstellen), soweit die betroffenen Teile über den Umfang üblicher Zitate hinausgehen.

3. Die ihre Bestände teilenden Quellenbanken können auch jene Quellen weitergeben, die sie selbst von einer anderen im Rahmen des allgemeinen Austausches empfangen haben.
4. Einzelnen Forschern, die sich der Dienste einer Quellenbank bedienen, erwächst die Verpflichtung, ihrerseits maschinenlesbar gemachte Quellen nach Abschluß der Auswertungen dieser Quellenbank zu übergeben.
5. Die einschlägigen infrastrukturellen Institutionen werden an die nationalen Finanzierungsträger der Forschung mit dem Vorschlag herantreten, in die Förderung aller datentechnisch unterstützten Projekte, bei denen maschinenlesbare Quellen entstehen, die explizite Verpflichtung der Projekte aufzunehmen, die entstandenen Daten kostenfrei nach Ende der Laufzeit des Projektes einer Quellenbank zur Verfügung zu stellen.
- f. Die konkrete Implementation einer Quellenbank im beschriebenen Sinn kann kaum Gegenstand einer über die jeweilige Institution hinausgehenden Vereinbarung sein. Von Interesse und einiger Bedeutung für die Verhinderung von Mehrfachentwicklungen wäre jedoch ein detaillierter Austausch über die Entwicklungsabsichten einschlägiger infrastruktureller Einrichtungen. Darüber hinausgehende konkrete Absprachen zur Arbeitsteilung wären wünschenswert. Dabei muß es primär um die Frage gehen, wieweit Designprinzipien für austauschbare Module in diesem Zusammenhang anzustreben sind und gemeinsame Prinzipien für die Softwareentwicklungen formuliert werden können.
- g. Große, unmittelbar nützliche und für die weitere Austauschbarkeit von Daten sehr zentrale Fortschritte können durch die Formulierung verbindlicher Regeln für die Dokumentation maschinenlesbarer Quellen, also jene Angaben gemacht werden, die in integrierte Quellenbanken als Bestandteile von Data Dictionaries auf einem hohen Niveau eingefügt werden müßten.

In diesem Zusammenhang ist zu überlegen, wieweit angestrebt werden sollte, im Rahmen einer systematischen Lesbarmachung großer Korpora, wie unter (c) erörtert, vorrangig die klassischen Quellenrepertorien aufzubereiten. Diese stünden der Disziplin damit in Hinkunft als durch Datenbanksysteme verwaltbare Einheiten zur Verfügung, es wäre also sehr viel einfacher, Änderungen und Ergänzungen einzuarbeiten.

- h. Mittelfristig von gleicher Bedeutung ist die Definition der bereits erwähnten Metafiles für einen Austausch von maschinenlesbaren Quellen zwischen einzelnen Quellenbanken, deren Inhalt dann durch geeignete Schnittstellen ohne menschliche Intervention in das lokale Datenbanksystem geladen werden kann. An besonderen Problemkreisen wären zu erwähnen:



- die einheitliche Beschreibung der für die Darstellung komplexer Zeichensätze verwendeten Konventionen.
  - die einheitliche Beschreibung der für die Darstellung struktureller Zusammenhänge zwischen einzelnen Quellenteilen verwendeten Konventionen; für die häufigsten Formen (etwa der hierarchischen Über-/Unterordnung) wäre in der Definition der Metafiles eine verbindliche Notation vorzusehen.
  - die einheitliche Beschreibung der Konventionen, die für die Darstellung formaler, nicht notwendigerweise struktureller Eigenschaften verwendet wurden. (Zu denken wäre etwa an die Verwendung kursiven Drucks für Ergänzungen in einer edierten Quelle, der in manchen Datenbanksystemen als niedrigere Sichtbarkeit der betroffenen Quellenteile zu implementieren ist, in statistisch ausgerichteten durch einen niedrigeren Wert einer GewichtungsvARIABLEN ausgedrückt wird und bei rein satztechnischen Systemen durch eine Übertragung in die entsprechenden lokal gültigen Druckkonventionen bearbeitet wird.)
- i. Einen wesentlichen Punkt bildet schließlich der Meinungsaustausch über die Richtungen, die in den einzelnen Organisationen für die Entwicklung von Software vorgesehen sind, um spezifische Auswertungen vorzunehmen, und die Frage, wieweit auch hier durch entsprechenden modularen Aufbau eine echte Arbeitsteilung durch die Kombinierbarkeit isoliert entwickelter Systeme gewährleistet werden kann.
  - j. Wichtig wäre schließlich auch ein Meinungsaustausch über verbindliche Vorstellungen einschlägiger Institutionen über die Leistungen, die akademische Rechenzentren in Zukunft erbringen sollten, um den voraussehbaren Interessen der Geisteswissenschaften zu entsprechen, insbesondere auch der Austausch der Vorstellungen, die jeweils über das künftige Verhältnis von Mikros und Mainframes gerade bei daten-/quellenbank-orientierten/ Projekten bestehen.

## **Möglichkeiten der Anregung zur Benutzung**

Wie wir bereits einleitend bemerkt haben, war eines der großen Probleme zentraler Datenarchivierung im Interesse der Bereitstellung von maschinenlesbaren Materialien für die Wiederverwertung, daß dem Angebot, dessen grundsätzliche Richtigkeit kaum in Frage gestellt wird, nur in relativ seltenen Fällen durch die Nachfrage entsprochen wurde: Einige der dafür verantwortlichen Ursachen haben wir bereits aufgezählt. Über das Gesagte hinaus scheint jedoch für eine erfolgsversprechende Installation wirklich verwendeter

Quellenbanken eine eingehendere Analyse der Problemlage angezeigt. Vor allem sollte man vier Gruppen potentieller Benutzer einer solchen Einrichtung unterscheiden:

- Forscher, die sich bereits bisher in ihrer jeweiligen Teildisziplin mit Erfolg datentechnischer Verfahren bedient haben.
- Forscher, die den Einsatz der Datentechnik zwar in Teilbereichen erwägen, durch die hohen Kosten der Vorbereitung der benötigten Quellen für die maschinelle Analyse aber davon abgeschreckt worden sind.
- Forscher, die an den einschlägigen Methoden und Techniken an sich zwar wenig oder nicht interessiert sind, sehr wohl aber an selektiven, auf einen bestimmten Quellenbereich bezogenen Dienstleistungen interessiert wären, wie sie durch den Einsatz der Mittel der Datentechnik möglich sind.
- Studenten, die zwar nicht in der Lage sind, aufwendige Projekte selbst zu initiieren, aber die methodischen Entwicklungen mit Interesse verfolgen und die Anwendung des in Veröffentlichungen beschriebenen Instrumentariums gerne in Form kleinerer, oft dissertationsbezogener Studien erproben möchten.

Um ein einschlägiges Angebot für alle vier Gruppen nutzbar zu machen, ist für jede dieser Zielgruppen eine eigene Vorgehensweise erforderlich.

Am einfachsten gestaltet sich das Problem sicherlich jenen Forschern gegenüber, die bereits bisher die Datentechnik eingesetzt haben. Das Angebot einer Quellenbank richtet sich ja auch am unmittelbarsten an diesen Kreis, dem dadurch die Möglichkeit geboten wird, intensiven Studien an Hand eines beschränkten, selbst aufbereiteten Quellenmaterials eine zusätzliche Dimension durch die flankierende Heranziehung weiterer Materialien zu verleihen. Auch hier ist allerdings die Information sicher noch zu verbessern, insbesondere durch die Verfügbarmachung detaillierter Beschreibungen des Inhaltes und der besonderen durch ihre jeweilige Kodierung gebotenen Hilfen. Praktisch wäre diese Aufgabe im Rahmen der bereits verschiedentlich als notwendig bezeichneten Quellenbeschreibung zu lösen.

Der zweiten Gruppe, also Forschern, die bisher an der Anwendung einschlägiger Verfahren durch die Kosten der Datenaufbereitung gehindert wurden, kann eine Quellenbank durch zwei Strategien entgegenkommen: wenn es gelingt, durch zentral zur Verfügung stehende Lesegeräte systematische Bemühungen um die Lesbarmachung größerer Korpora zu erreichen, wird deren Effizienz in entscheidendem Maße davon abhängen, wieweit gerade die eben angesprochene Gruppe hinreichend von den dadurch gebotenen Möglichkeiten informiert werden kann. Unabhängig davon muß erreicht werden, daß einschlägige Institutionen, die bereits bisher methodische/technische

Beratung geleistet haben, in die Lage versetzt werden, gegebenenfalls darauf hinzuweisen, daß Teile der angesprochenen Fragestellungen mit bereits verfügbarem Quellenmaterial bearbeitet werden können. Natürlich kann es nicht darum gehen, die Forschung auf jene Quellen zu spezialisieren, die – von der Warte des mit dem Angebot Konfrontierten mehr oder minder zufällig – gerade bearbeitbar vorliegen. Wir glauben aber, daß in allen jenen Fällen, wo der Wert des Einsatzes einschlägiger Verfahren zweifelhaft ist, die Erprobung der geplanten Techniken an Hand nicht allzu unterschiedlichen Materials eine wertvolle Entscheidungshilfe geben könnte. Wovon letztlich auch sehr viele Projekte profitieren könnten, die derzeit den Einsatz der Datenverarbeitung sozusagen "blind" planen, ohne sich der Möglichkeit zu Pretests vor Beginn der aufwendigen Datenerfassung bewußt zu sein.

Als Dienstleistung für selbst nicht an formalen Analysen arbeitenden Kollegen ist zu prüfen, wieweit die an einer Institution vorgesehenen spezifischen Auswertungen geeignet sind, routinemäßig in vereinfachter Form auf wesentliche Teile des zur Verfügung stehenden Gesamtmaterials angewendet zu werden. Möglich wird dies freilich nur sein, wenn man sich von der Vorstellung frei macht, in jedem Fall perfekte, auf eine Lebensdauer von Jahrzehnten berechnete Behelfe zu liefern und statt dessen beispielsweise die Möglichkeit vorsieht, rein mechanisch Hilfsmittel eingestandenermaßen niedriger Qualität einmal zu erstellen, auf Microfiche zu speichern und nur auf Anforderung zugänglich zu machen.

Vielleicht der wichtigste Punkt ist der letzte: die Schaffung von Möglichkeiten für Studenten, sich bereits zu einem frühen Zeitpunkt mit den neu entstehenden Techniken und Methoden vertraut zu machen, ohne daß dies in jedem Fall mit größeren Hintergrundinvestitionen verbunden werden muß. Zu diesem Zweck wäre eine enge Kooperation mit den Universitäten anzustreben. Konkret sollte einerseits die Möglichkeit ins Auge gefaßt werden, interessierten Universitätslehrern geeignete maschinenlesbare Quellen zur exemplarischen Analyse im Rahmen einer Lehrveranstaltung zur Verfügung zu stellen, andererseits sollte auch immer wieder mit Nachdruck betont werden, welche Möglichkeiten sich für komparatistische Ansätze durch die Verfügbarkeit großer maschinenlesbarer Korpora ergeben.

Der letzte Punkt scheint uns so wichtig, daß wir ihn auch gesondert nochmals hervorheben möchten: eine der wesentlich neuen Möglichkeiten, die durch die Verfügbarkeit maschinell verarbeitbaren Materials gegeben sind, liegt jedenfalls in der Verbesserung der Ausgangsbedingungen komparatistisch verstandener Ansätze. U.E. sollten Quellenbanken gezielt versuchen, interessierte Forscher von sich aus darauf hinzuweisen, daß für sie relevante Materialien vorliegen, deren Bearbeitung in der Zusammenschau Ergebnisse

erbringen kann, die über die Summe der von getrennt voneinander arbeitenden Einzelstudien erbrachten deutlich hinausgehen.

### **International Ressource Sharing.**

Wir haben bereits betont, daß die hier vorgeschlagenen Ansätze Teil einer Infrastruktur bilden, die, um effektiv zu sein, über nationale Grenzen hinweg wirken muß.

Unmittelbare Konsequenz dessen war bereits die Forderung, daß alles maschinenlesbar gemachte Material grundsätzlich der ganzen wissenschaftlichen Gemeinschaft ohne weitere Kosten zur Verfügung stehen sollte. Mittelbar bedeutet ein derartiger Grundsatz natürlich, daß in einem Land getätigte Investitionen – etwa ein gezielt eingesetztes Lesegerät – grundsätzlich auch Projekten zugute kommen, die jenseits der Landesgrenzen beheimatet sind.

Wir glauben, daß dies nicht ein bloßer Zufall sein sollte, der sich aus der Natur von Quellenbanken ergibt, sondern daß die Realisierung und Bewältigung dieser Situation eine wesentliche Vorbedingung für Fortschritte bei der Nutzbarmachung der neuen Informationstechnologien für das Fach schlechthin bedeutet. Die Anwendung neuer Informationstechnologien im Bereich der Geschichtswissenschaften wird sicher stets zu einem hohen Grad von der Übernahme in anderen Bereichen entwickelter Problemlösungen getragen werden; andererseits scheint es uns zweifelsfrei, daß aus den allgemeinen Voraussetzungen historischer Forschung letztlich eine ganze Reihe von Anforderungen an eine Methodologie der Informationserschließung historischen Quellenmaterials abgeleitet werden kann und muß, die von in anderen Bereichen geleisteten Entwicklungen nie abgedeckt werden können. Jetzt schon läßt sich absehen, daß die dabei zu bewältigenden Schwierigkeiten über das Potential ad hoc geschaffener Einzellösungen weit hinausgehen.

Allerdings ist angesichts der heutigen Kosten professioneller Programmentwicklung durch kommerzielle Häuser und der vergleichsweise geringen Zahl von Historikern im allgemeinen – ganz zu schweigen von den wenigen, die jetzt schon die Datenverarbeitung in anspruchsvoller Form heranziehen – kaum anzunehmen, daß Auftragsentwicklungen größeren Maßstabes möglich sind. Zur Bereinigung dieser Situation sind u.E. zwei Strategien möglich:

- einschlägige Entwicklungen in infrastrukturellen Einrichtungen und längerfristig arbeitenden Forschungsprojekten sollten von vornherein soweit abgestimmt werden, daß sowohl die verwendeten Datenformate austauschbar als auch die einzelnen Programmsysteme so modular und damit übertragbar wie möglich gehalten werden.
- es muß die Möglichkeit geschaffen werden, zur Erprobung neuer Ansätze (in klarerweise beschränktem Ausmaß) Ressourcen der Institutionen, die

an entsprechenden Entwicklungen arbeiten, auch durch auswärtige Projekte mit zu benutzen. Dies kann nicht heißen, daß Forschungsaufgaben einer mit einem unzureichenden Rechenzentrum ausgestatteten Universität stillschweigend ins besser dotierte Ausland verlagert werden. Es soll und muß unseres Erachtens aber möglich sein, methodisch interessante Pilotstudien zur Erprobung neuer Methoden und Techniken an allen in der methodischen Entwicklung stehenden Institutionen auch dann durchzuführen, wenn dadurch Projekte eines anderen, auch dem jeweiligen Ausland angehörigen Finanzträgers in substantieller Weise gefördert werden, ohne in jedem einzelnen Fall zeitraubende Überlegungen über die Übertragung eingesetzter Fördermittel über die Grenzen hinweg anzustellen.

## Überlegungen zu einer Datenbank "Historische Demographie"

Am Beispiel der historischen Demographie soll in den folgenden Überlegungen der Versuch gemacht werden zu zeigen, welche Informationen der Benutzer einer Datenbank 'Historische Statistik' erwarten können sollte, welche Quellengruppen für einen solchen Zweck in Frage kämen und in welcher Form die aus ihnen gewonnenen Daten zugänglich gemacht werden müssten. Dieser Problemkatalog ist sehr umfanglich. Ihn vollständig abzudecken, ist in diesem Beitrag nicht möglich, nicht einmal für den gesamten Bereich der historischen Demographie. Außerhalb der Betrachtung bleiben darüberhinaus weitgehend Fragen des formalen Aufbaus von Datenbanken.

Die wichtigste Frage bei der Einrichtung einer Datenbank 'Historische Statistik' ist die nach den präsumtiven Benutzern - ihrer Vertrautheit im Umgang mit einem Rechner und ihren möglichen Wünschen an Information wie Komfort. Es liegt auf der Hand, daß der Historiker, der selbst an Forschungsproblemen sitzt, die den Rechnereinsatz erfordern, weniger an Komfort als an einem möglichst vielseitigen Instrument für seine Auswertungen interessiert ist. Hat sich der Forscher dagegen in einen Lehrer verwandelt, so wird er eher an bequem abrufbarer Information - möglichst bereits statistisch und in Tabellen- bzw. graphischer Form aufbereitet - Gefallen finden. Da Datenbanken ein recht rigides Design erfordern, das Ergänzungen und Erweiterungen nur zum Teil unproblematisch zuläßt, da Datenbankkonstrukteure, was die Wünsche von präsumtiven Benutzern angeht, entweder gut zu halluzinieren imstande sein oder die gängigen Fragestellungen als Raster benutzen müssen, ist für die erste Art von Benutzern eine Datenbank nur von begrenztem Nutzen. Ihnen wird ein herkömmliches Dateienkonzept vermutlich reichen. Die eigentliche Zielgruppe ist die der Lehrenden im weitesten Sinne. Sie mit forschungsnahen Informationen zu versorgen, wäre die wichtigste Aufgabe einer Datenbank. Aufgabe der Forschung wäre es, Daten zu liefern: Welche und in welcher Form, das soll am Beispiel der historischen Demographie erläutert werden.

Der neuesten Arbeit von *Dupaquier* ist zu entnehmen, daß in Frankreich 538 Ortsmonographien entstanden sind, die sich in sehr zufälliger Weise über die französischen Departements verteilen. Hier geht es nicht darum, die Folgen einer fehlenden Forschungsstrategie für die Repräsentativität der Einzelergebnisse abzuschätzen (*Dupaquier*, *Demographie*, 110f), sondern darum, daß ein großer Teil der in den genannten Monographien enthaltenen Informationen schwer zugänglich ist. Selbst wenn Repräsentativität und Methodik dem einen oder anderen Zweifel unterliegen mögen, nicht zweifeln läßt sich

an der Notwendigkeit, die gewonnenen Ergebnisse anderen in einer näher zu spezifizierenden Form verfügbar zu machen. Welche ungenutzten oder nur einseitig genutzten Datenfriedhöfe anderswo noch vorhanden sind, kann nur vermutet werden. Daß es solche in großem Umfang und nicht nur im Bereich der historischen Demographie gibt, ist mit Sicherheit anzunehmen.

Nicht umsonst sind allerdings solche Datenfriedhöfe entstanden: das Material zu ordnen und damit erst verfügbar zu machen, das ist mit einem erheblichen, kaum zu unterschätzenden Aufwand verbunden. Nach Abschluß einer Untersuchung ist es keinem zuzumuten, seine Daten in eine auch anderen zugängliche Form zu bringen, zumal Kosten dadurch entstehen und zumindest die wichtigsten Ergebnisse ja sowieso veröffentlicht werden. Die Kosten für den Aufbau einer Datenbank 'Historische Statistik' rechtfertigen sich daher nur durch einen entsprechend großen Abnehmerkreis. Universitäre Forschung ist kein ausreichender Grund.

Die tautologische Aussage, eine Datenbank 'Historische Demographie' solle ausschließlich aus der Vergangenheit gewonnene Bevölkerungsdaten enthalten, gewinnt an Aussagekraft mit dem Hinweis auf das dichte und vielfältige Geflecht sozioökonomischer Tatbestände, in das die Bevölkerungsdaten eingebettet sind. An dieser Stelle wird deutlich, daß eine Datenbank 'Historische Demographie' allenfalls sinnvoll erscheint innerhalb einer umfassenden Datenbank 'Historische Statistik'. Zahlreiche Aussagen der historischen Demographie - gerade auch ein großer Teil der interpretatorisch relevanten - sind nur möglich in Verbindung mit weiteren sozioökonomischen Variablen: herrschaftliche Zugehörigkeit, Erbrecht, Vererbungsgewohnheiten, Zahl und Art der wirtschaftlichen Betriebe, Betriebsgrößenstruktur, sektorale bzw. branchenspezifische Zugehörigkeit, Erträge aus landwirtschaftlicher und/oder gewerblicher Tätigkeit, Preise, Löhne usw. Ein Teil solcher Daten wird aus klassischen demographischen Quellen gewonnen. Doch ist das nicht der Normalfall. Hier wird zu überlegen sein, wo und wie solche Daten abgelegt werden. Bevor aber über die Zuordnung solcher sozioökonomischer Daten zu demographischen etwas gesagt werden kann, müssen letztere klassifiziert werden, um entscheiden zu können, welche davon und in welcher Form in eine Datenbank 'Historische Demographie' aufgenommen werden sollten.

Eine vollständige Übersicht über bevölkerungsgeschichtliche Quellen, über deren unterschiedlichen formalen Aufbau sowie deren inhaltliche Qualitäten und über damit verbundene Auswertungsmethoden gibt es nicht, für den Bereich der deutschen Archive schon gar nicht. *Willigan und Lynch* bieten allerdings eine brauchbare Übersicht (*Willigan/Lynch, Sources*). Darin fehlt freilich ein Hinweis auf die nicht wenigen Quellen, die bereits ab dem 17. Jahrhundert - zuweilen auch bereits früher - recht differenzierte nicht

nominative und/oder aggregierte Bevölkerungsdaten enthalten. Ich denke hier z.B. an Populationstabellen der badischen und fürstenbergischen Territorien, die ab ca. 1750 teilweise Jahr für Jahr angelegt worden sind, oder an andere Einwohnerlisten: Herdstätten- oder Häuserzählungen. Grundsätzlich sind also von den Quellen her bereits nicht nominative und/oder aggregierte Daten von solchen zu unterscheiden, in denen Individuen genannt sind. Bei den aggregierten Daten sind im übrigen die den Quellen so entnommenen von den erschlossenen zu unterscheiden, wobei eine solche Differenzierung keineswegs eine solche der Qualität ist. Das gilt freilich weniger für summarische Kommunikanten- und Seelenzahlen; da sie eine häufig nicht ersetzbare Überblicksinformation bieten, sind solche Daten dennoch wertvoll. Kirchenbücher und Haushaltslisten sind an vielen Stellen beschrieben worden (zuletzt bei *Willigan/Lynch, Sources*, 57-104). Auf sie muß daher nicht eingegangen werden. Die Zahl der möglichen Quellen ist damit abgeschritten. Dabei sind Quellen außer acht gelassen worden, die auch, aber nicht in erster Linie für bevölkerungshistorische Zwecke benutzt worden sind: z.B. sämtliche Quellen zur Besteuerung (*Willigan/Lynch, Sources*, 133-159).

Die Frage, welche demographischen Maßzahlen berücksichtigt werden sollen, führt in das Zentrum der Problematik beim Aufbau einer Datenbank 'Historische Demographie'. Sind schon die Quellen nicht ausreichend geordnet und klassifiziert, so finden wir in weiten Bereichen der historischen Sozialwissenschaft ungelöste Definitions- und Berechnungsprobleme - es sei hier nur an die Arbeiten zur sozialen Schichtung erinnert - , nach deren Lösung erst Daten sehr unterschiedlicher örtlicher und zeitlicher Provenienz vergleichbar werden. Glücklicherweise sind die meisten Arbeiten zur historischen Demographie, insbesondere die auf Rekonstruktionsmethoden sich stützenden, nach 1954 erschienen und haben die methodischen Vorschläge *Henrys* berücksichtigt. Ein großer Teil der demographischen Daten für den Zeitraum vor 1850 ist durch Kirchenbuchanalysen erarbeitet worden. Auf Probleme der Datenaufbereitung aus solchen Analysen soll zunächst eingegangen werden. Anschließend sind Daten zu behandeln, die aus sogenannten *status animarum* gewonnen sind: darunter sind hier sämtliche Quellen begriffen, in denen auf Individuen bezogene, aber zu Familien/Haushalten zusammengefaßte Daten für einen bestimmten Zeitraum enthalten sind. Schließlich ist von der sehr vielfältigen Gruppe nicht nominativer Daten, freilich nur am Beispiel der Populationstabellen zu sprechen.



## 1 Kirchenbücher

Aus Kirchenbüchern erhobene Daten weisen bis auf wenige Ausnahmen zwei Besonderheiten auf: a. die klassischen Maßzahlen der Demographie - Geburten-, Heirats-, Todesrate - können nicht berechnet werden, da Angaben zur Gesamtbevölkerung fehlen; b. meistens ist die Zahl der verarbeiteten Fälle so gering, daß allenfalls Methoden deskriptiver Art in Frage kommen, da analytische Methoden Normal- und andere normierte Häufigkeitsverteilungen berücksichtigen und damit entsprechend hohe Fallzahlen voraussetzen, von Stichprobenanalysen einmal abgesehen. Scheingenauere Berechnungsergebnisse mit scheinbaren Methoden aus unzureichenden Daten führen nur in die Irre. Gleiches gilt für Versuche der Schätzung von Maßzahlen aus Rekonstitutionsdaten (*Willigan/Lynch, Sources, 182*), die dem Verdacht ausgesetzt sind, aus zuweilen schlechten Zahlen noch schlechtere, weil zu wenig differenzierte und aussagekräftige Schätzungen zu machen.

Weitgehend unproblematisch - weil die Rohdaten lediglich beschreibend und ordnend, nicht für weitere statistische Analysen verwendend - erscheint die Aufnahme vitalstatistischer und aus Familienrekonstitutionen gewonnener Daten in eine Datenbank.

1.1 Vitalstatistische Reihen bieten angesichts der geringeren Datenqualität nur wenige Möglichkeiten einer zumeist recht groben Orientierung:

1.1.1 Das Verhältnis von *Taufen* zu *Todesfällen* kann einen ersten Eindruck vermitteln von den langfristigen Trends, d.h. darüber, ob eine Bevölkerungszahl gleich bleibt, kleiner oder größer wird. Vorausgesetzt ist, daß mögliche Datenfehler - gedacht ist an solche, die sich z.B. bei schlechter Führung der Kirchenbücher ergeben - bereits korrigiert und mit entsprechender Gewichtung die zugrundeliegenden Zahlen geändert sind (*Wrigley/ Schofield, Population History, 15-154, bes. 174-189*). Die Tabellen sollten gleitende Durchschnitte der vitalstatistischen Zeitreihen sowie des Quotienten bieten. Denkbar wäre eine Kumulierung zu fünfjährigen Intervallen. Sinnvoll wäre es, kumulierte Differenzen zu berechnen, weil so in der bildlichen Darstellung eine Verminderung oder Vermehrung der Bevölkerung sichtbar gemacht werden kann.

1.1.2 Das Verhältnis von *Taufen* zu *Eheschließungen* steht methodisch und daher vom möglichen Aussagewert her auf ebenso wackligen Beinen. Darauf ist bereits hingewiesen worden (*Henry, Techniques, 48; Wrigley / Schofield, Population History, 189f.*). Die Daten könnten in der eben beschriebenen Art aufbereitet werden.

Soweit die vitalstatistischen Daten rekonstruiertem Material entstammen sind drei weitere Zahlenreihen zu erstellen:

- 1.1.3 Die *saisonale Verteilung* von Geburten/Konzeptionen, Eheschließungen und Todesfällen über das Jahr *Wrigley/Schofield*, Population History, 286-305) sollte in 25- bis 50- Jahresintervallen in einer Tabelle erfaßt werden.
- 1.1.4 Die *Geschlechterproportion* bei Taufe/Geburt und Tod dient vor allem der Kontrolle, da sie keinen weitergehenden Aussagewert besitzt (*Henry*, Manuel, 11; *Henry*, Techniques, 46f.). Es soll trotz dieser Geringerbewertung nicht übersehen werden, daß für die Beurteilung von Bevölkerungen gerade bei Abweichungen vom normalen Intervall sich aus der Geschlechterproportion wichtige Hinweise für die weitere Analyse ergeben.
- 1.1.5 Die *Illegitimitätsquote* sollte ebenfalls aufgenommen werden, obwohl vor 1800 wie bei den Zwillingsgeburten die Zahlenwerte so gering sind, daß der Informationswert zumindest fraglich ist.

So mager der Ertrag aus der Verarbeitung vitalstatistischer Daten sein mag, deren Aufnahme in eine Datenbank rechtfertigt sich aus folgenden Gründen: In der Literatur, auch der nichtprofessionellen ortsgeschichtlichen, sind solche Zahlenreihen häufig angegeben. Diese sind - soweit um Registrationsfehler bereinigt - problemlos vergleichbar und kumulierbar. Nicht zuletzt für die Forschung, die mit ausgefeilteren Methoden aus solchen Daten einiges herausholen kann, sind diese Zahlen von einigem Wert (*Livi Bacci*, Demographic Trends; *Lee*, Methods and Models).

- 1.2 *Rekonstruktionsmaterial* bietet dagegen differenzierte und verlässlichere Maßzahlen, allerdings für die Zeit vor 1850 zumeist generationsspezifisch berechnet. Dadurch verlieren solche Zahlen - vor allem im zeitlichen und regionalen Vergleich - gegenüber periodenspezifisch berechneten keineswegs an Aussagekraft. Diese leidet vielmehr dadurch, daß die Zahl der sicher und vollständig rekonstruierten Familien eines Ortes zuweilen nur einen geringen Bruchteil der insgesamt erfaßten Personen umfassen kann (*Sachs*, Bevölkerung, 151). Folgende Informationen sollten bei einer Datenbank abrufbar sein:
  - 1.2.1 Das *Heiratsalter* ist zu differenzieren nach Erst- und Mehrfachehen. Zu welchen Zeiträumen diese Zahlen kumuliert werden sollen, kann nicht für alle Fälle normiert werden. Ein fünfjähriger Zeitraum sollte angesichts der meist geringen Zahl der Ehen der Normalfall sein, zumal auch im nachhinein diese Zahlen problemlos zu größeren Zeiträumen kumuliert werden können. Wie auch für die im folgenden genannten Zahlenreihen ist im übrigen immer die Zahl der zugrundeliegenden Fälle, möglichst auch ihr Anteil an der Gesamtzahl anzugeben, eine Übung, die in der Literatur bisher zu wenig beachtet worden ist.

- 1.2.2 *Natalitäts-* sowie die dazugehörigen Reproduktionsziffern lassen sich generationsweise berechnen. Reproduktionsziffern wären allerdings nur dann sinnvoll, wenn bei Vorliegen einer Sterbetafel Nettoreproduktionsziffern in einer Tabelle dargestellt werden könnten (*Esenwein - Rothe*, Einführung, 322f.).
- 1.2.3 Für die *Fertilitätsziffern* bieten sich die von *Henry* vorgeschlagenen Berechnungsmethoden an (*Henry*, Manuel 79-80; *Henry*, Techniques, 70-84). Die Tabellen sollten die Fertilitätsziffern nach Alter der Frauen und Alter bei der Heirat enthalten. Bei den Altersangaben sind Fünfjahresintervalle sicherlich sinnvoll (insgesamt sechs Intervalle auf die proliferationsfähige Zeit). Für die Größe der Zeiträume bzw. der Generationen kann auch hier keine verbindliche Regel gegeben werden. Ob als weiteres Differenzierungsmerkmal Beruf bzw. soziale Position berücksichtigt werden können, hängt von der Datenqualität, mehr noch aber von den Möglichkeiten ab, ein regional und zeitlich übergreifendes Einordnungsraster zu finden.
- 1.2.4 *Proto- und intergenetische Intervalle* können ebenso wie Fertilitätsziffern nach Lebensalter und Ehedauer der Frauen tabellarisch erfaßt werden (*Henry*, Manuel, 97-102; *Henry*, Techniques, 102-109). Ob nach Alter der Frau und nach Rang geordnete intergenetische Intervalle berechnet werden, muß angesichts der Berechnung altersspezifischer Fertilitätsziffern und mit Blick auf das Verhältnis zwischen Aufwand und Aussageertrag reiflich überlegt werden.
- 1.2.5 *Mortalitätsziffern* können nach der von *Henry* zur Berechnung der Kindersterblichkeit vorgeschlagenen Methode tabellarisch aufbereitet werden (*Henry*, Manuel, 122f.). Dabei geht es um die Berechnung der Sterbewahrscheinlichkeit. Voraussetzung sind verlässliche Angaben zum Alter bei Eintritt des Todes. Grundsätzlich sollten nur Personen in diesem Datensatz enthalten sein, deren Geburts- und Todesdatum eindeutig feststeht. Unter diesen Voraussetzungen läßt sich generationen-, d.h. zeitraumspezifisch eine Abfolge von *Sterbetafeln* erstellen, die neben der Sterbewahrscheinlichkeit ( $q$ ) noch die Ziffern zur durchschnittlichen Lebenserwartung ( $e$ ), vielleicht sogar die von *Mackenroth* erwähnte wahrscheinliche Lebenserwartung enthalten. (*Mackenroth*, Bevölkerungslehre, 80f.).
- 1.2.6 Die *Kinderzahl* je abgeschlossene Ehe sollte für bestimmte Zeiträume, u.U. differenziert nach weiteren Merkmalen, gesondert, d.h. neben Ehestands- und Natalitätsziffern erfaßt sein, um problemlose Vergleiche mit anderen Quellengruppen zu ermöglichen.
- 1.2.7 Ob neben dem Heiratsalter noch weitere Ziffern zur *Nuptialität* erhoben

werden sollten, wird hier offengelassen. Sinnvoll wäre es sicherlich, den Personenstand der Heiratenden zu berücksichtigen. Die Herkunftsorte der Eheleute wären in einer sozialgeographischen Studie zum Heiratsverhalten - z.B. zur Berechnung eines örtlich spezifischen Heiratsradius - verwendbar, ebenso wie die soziale Herkunft einen breit gefächerten sozialhistorischen Vergleich ermöglichen würde. Letzteres steht allerdings unter den in 1.1.3 bereits genannten Vorbehalten.

## 2 Seelenbeschriebe

Seelenbeschriebe vermögen besonders in Hinsicht auf Familien/ Haushaltszusammensetzung und Altersaufbau zusätzliche Informationen zu geben. Aus ihnen können außerdem Sterbetafeln nach dem Periodentafelkonzept abgeleitet werden. Alle diese Daten sollten entsprechend dem bisherigen Organisationsprinzip nach geographischen und zeitlichen Räumen geordnet sein. Im übrigen lassen sich hier leichter Merkmale der sozialen Schichtung mit bevölkerungs- und familienhistorischen Merkmalen verknüpfen.

- 2.1 Seelenbeschriebe gibt es in unterschiedlicher Qualität. Danach richten sich die Möglichkeiten der Auswertung. Aus Seelenbeschrieben, in denen nur Eltern und Kinder genannt sind, lassen sich lediglich durchschnittliche *Familien/ Haushaltsgröße, Geschlechterproportion, Zahl der Kinder je Familie/Haushalt* erheben.
- 2.2 Diese Daten können je nach Qualität des Seelenbeschriebs differenziert und erweitert werden.
  - 2.2.1 Sind *Beruf* und *Zivilstand* der Haushaltsmitglieder angegeben, so kann die jeweilige *Kinder-, Gesinde- und Inwohnerzahl* differenziert in Hinsicht auf die soziale Gliederung berechnet werden.
  - 2.2.2 *Altersangaben* - leider nicht sehr häufig anzutreffen und bisweilen auch nur für die männliche Bevölkerung und die über Vierzehnjährigen - erlauben eine Reihe weiterer Analysen: *Altersgliederung, Sterbetafel* nach dem Periodenkonzept mit Sterbewahrscheinlichkeit und *Lebenserwartung, Kinder- und Gesindezahl* in Abhängigkeit vom *Alter des Familienvorstandes*.
  - 2.2.3 In manchen Seelenbeschrieben finden sich *Nachträge* und *Ergänzungen*; es werden z.B. aus dem Haushalt durch Heirat, Tod oder Stellenwechsel ausgeschiedene Personen gestrichen bzw. das entsprechende Datum vermerkt (*Hesse, Ganzes Haus*, 137-139). Sind diese Angaben vollständig, so ließen sich Aussagen zu *Fluktuation* sowie *Familien/Haushaltszyklus* machen.

- 2.3 Sind Seelenbeschriebe in genügender Dichte - d.h. der Abstand sollte vor 1750, bzw. bei nur langsam sich verändernden Bevölkerungen, nicht zehn, nach 1750, bzw. bei schnell oder sprunghaft sich verändernden Bevölkerungen, nicht fünf Jahre überschreiten - vorhanden, ließen sich vorsichtige Aussagen zur *Migration* machen. Das wäre allerdings nur im Zusammenhang mit Kirchenbüchern möglich, die eine Abzählung der natürlichen Zu- und Abgänge (Geburten, Todesfälle) erlauben.

### 3 Populationstabellen

Populationstabellen - entstanden z.B. in den badischen und fürstenbergischen Territorien im Zuge physiokratisch inspirierter Bevölkerungs- und Wirtschaftspolitik - enthalten Informationen zu einzelnen Orten, zu Ämtern, zu ganzen Territorien. Im allgemeinen sind genannt: Gesamtzahl der anwesenden - zuweilen auch abwesenden - Personen nach Geschlecht und jeweils nach großen Altersgruppen gesondert aufgeführt (z.B.: unter 15, 15-50, über 50 Jahre), Zahl der Taufen/Geburten, der Eheschließungen und der Begräbnisse/Sterbefälle im Kalenderjahr. Bei einigen Erhebungen sind weitere Informationen zu Stand, Beruf, Familien- oder Häuserzahl verzeichnet.

- 3.1 Auf der Basis dieser Quellen lassen sich auf die Gesamtbevölkerung bezogene *Mortalitäts-, Natalitäts- und Nuptialitätsziffern* berechnen und zwar ggf. in langfristigen Zeitreihen.
- 3.2 Die Entwicklung der *Altersgliederung* kann Ort für Ort wie auch geographisch kumuliert erfaßt werden. Allerdings handelt es sich nur um sehr grobmaschige Angaben.
- 3.3 Liegen Populationstabellen in dichter Folge vor, können Zahlen zur *Migration* berechnet werden.

Es führt zu weit, sämtliche Möglichkeiten der Auswertung dieser Quellengruppe aufzuführen. Deutlich ist aber, daß der Gehalt solcher Quellen, in der angedeuteten Form tabellarisch erschlossen, als Überblicksinformation sich ganz hervorragend für die Zwecke einer Datenbank eignen würde.

In meinen bisherigen Überlegungen sind einige wichtige Quellengruppen, ihr Informationswert und einige der Verarbeitungsmöglichkeiten zur Sprache gekommen. Ausgegangen wurde dabei von Benutzern, die an Zwischen- und Endergebnissen der Verarbeitung allein interessiert sind, möglichst in Tabellenform bereits aufbereitet und über kleinere Benutzerprogramme am Bildschirm in einer graphischen Darstellung sichtbar zu machen. Bereits in dieser Form wäre auch Interessen der reinen Forschung gedient. Klar ist aber, daß für den Forschungsbereich auch die Rohdaten zur Verfügung stehen müssen.

Es fragt sich, in welcher Form.

Grundsätzlich könnten die Rohdaten in der herkömmlichen Dateienform (Zeilen/Satz - , Spaltenaufteilung mit beliebigen Längen) Teil einer Datenbank werden. Eine solche Entscheidung liegt schon deswegen nahe, weil wohl die meisten bisher erhobenen Daten zur historischen Demographie in eben dieser Form bereitliegen. Eine Umstellung würde zu erheblichen Kosten führen. Da Daten der historischen Demographie aber keine sehr große Variationsbreite aufweisen, dürfte es genügen, wenn der Benutzer am Anfang einer Fremddatei eine inhaltliche und syntaktische Beschreibung vorfindet, die ihn instandsetzt, auf Einzeldaten zuzugreifen. Beigegeben sein könnten für die Auswertung der Daten bereits erfolgreich benutzte Programme.

Wenig Probleme bereiten in dieser Form die unter (1.1), (2) und (3) behandelten Quellengruppen. Für die Kirchenbuchauswertungen, die bis zur Familienrekonstitution gekommen sind, empfiehlt sich eine Teilung der Daten in sicher und nicht rekonstituierte. Erstere sollten in der Art eines Familienbuches oder gemeinsam mit einem solchen sämtliche sicher erhobenen Daten enthalten. Das übrige Namensmaterial würde dem Benutzer an anderem Ort zur Verfügung gestellt werden.

Dieses Namensmaterial ist insofern wichtig, als neben Ortsnamen nur über die Eigennamen die Verbindung zu anderen Datenserien hergestellt werden kann: (Grund)besitz und Besteuerung, freiwillige und Strafergerichtsbarkeit, Vertragsschlüsse aller Art, z.B. Versicherungen. Diese Liste könnte beliebig verlängert werden. Auch ohne das wird deutlich, daß die inhaltlichen Grenzen einer Datenbank 'Historische Statistik' damit bereits überschritten werden. Weiterführende Überlegungen zur Verknüpfung der Daten sind hier nicht möglich, da genaue Vorstellungen über das Datenbankkonzept und über die formale Gestaltung der Datenbank voraussetzte.

Es kann hier auch nicht mehr auf Probleme der Beschreibung von Datensätzen eingegangen werden, nicht zuletzt weil damit gleichzeitig Probleme der Datenaufnahme, ihrer Normierung und Verarbeitung verbunden sind, Probleme, die schwierig zu lösen sind, weil entweder koexistierende Programmpakete vereinbar gehalten oder aber verbindliche Normen gefunden werden müßten. An Schärfe verliert aber die Problematik dadurch, daß sie sich ausschließlich für den Forschungsbereich stellt, für den eine gewisse Vertrautheit mit datentechnischen Fragen oder Datentechnikern vorausgesetzt werden kann.

Abschließend sei noch kurz auf Überlegungen eingegangen, die *Johansen* vor einigen Jahren angestellt hat hinsichtlich der Errichtung demographischer Datenbanken (*Johansen, Priority Problems*). Er ist von der Voraussetzung ausgegangen, daß Datenaufnahme und -verarbeitung Sache der Datenbankverwalter sei. Unter einer solchen Voraussetzung werden die Kosten für eine

Datenbank hochgetrieben. Es scheint mir aber wirklichkeitsnäher, wenn davon ausgegangen wird, daß die Datenbankverwalter tatsächlich nur verwalten: Normierung der Verarbeitungsschritte, Organisation der Daten, Erstellung von Benutzerrouitinen. Die Datenerhebung und -verarbeitung bleibt in einem solchen Konzept Sache der Forschung.

### Literaturverzeichnis

Appel, E. /Karlsson, K.H. / Kronborg, B.: Computerized family reconstitution. In: Raben, J. / Marks, G., Data bases in the Humanities and the Social Sciences, Amsterdam, 1980, 125-129.

Beauchamp, P. / Charbonneau, H. / Dejardins, B. / Legare, J. : La reconstitution automatique des familles. Un fait acquis, in: Population Num. spec. 1977, 375-399.

Bortkiewicz, L. v. : Die Bevölkerungstheorie. In: Die Entwicklung der deutschen Volkswirtschaftslehre in neunzehnten Jahrhundert. Gustav Schmoller zur siebenzigsten Wiederkehr seines Geburtstages (...) Erster Theil, Leipzig, 1908, XIII, 1-57.

Bouchard, G. / Pouyez, C.: Name variations and computerized linkage, in: HMN 13/2 (1980), 119-125.

Chamoux, A.:La reconstitution des familles. Espoirs et réalités, in: Annales E.S.C. 27 (1972), 1083-1090.

Charbonneau, H. / Legare, J.: Utilisation des ordinateurs en démographie historique. In: Marcilio, M.L. /Charbonneau, H., Démographie, Rouen / Montreal, 1979, 109-130.

Daubeze, Y. / Perrot, J.-C.: Un programme d'étude démographique sur ordinateur, in: Annales E.S.C. 27 (1972), 1047-1070.

Desjardins, B. / Beauchamp, P. / Legare, J.: Automatic family reconstitution: The French-Canadian seventeenth century experience. In: JFH 2 (1977), 56-76.

Dupaquier, J.: Pour la démographie historique, Paris, 1984.

Esenwein-Rothe, I.: Einführung in die Demographie. Bevölkerungsstruktur und Bevölkerungsprozeß aus der Sicht der Statistik (=Stat. Stud. 10), Wiesbaden, 1982.

Gutmann, M.P.: Reconstituting Wandre. An approach to semi- automatic family reconstitution, in: ADH 20 (1977), 315-341.

Henry, L.: Manuel de démographie historique, Genève/Paris, 1972.

Henry, L.: Techniques d'analyse en démographie historique, Paris, 1980.

Hesse, G.: Ganzes Haus und Kernfamilie (...), in: SVGB 102 (1984), 130-139.

Johansen, H.C.: Priority Problems in establishing demographic data bases, In: Sundin, J. / Soederlund, E.: Time, Stockholm, 1979, 239-243.

- Lee, R.D.: Methods and models for analysing historical series of births, deaths, and marriages, In: *ders.*: Population patterns, New York, 1977, 337-370.
- Lee, R.D.: Models of preindustrial population dynamics with application to England, In: Tilly, C.: Historical Studies, Princeton, 1978, 155-207.
- Lee, R.D.: Population patterns in the past, New York, 1977.
- Livi Bacci, M.: Can anything be said about demographic trends when only aggregate vital statistics are available? In: Lee: Population, New York, 1977, 311-336.
- Mackenroth, G.: Bevölkerungslehre. Theorie, Soziologie und Statistik der Bevölkerung, Berlin / Göttingen / Heidelberg, 1953.
- Marcilio, M.L. / Charbonneau, H. (eds.): Démographie historique, Rouen / Montreal, 1979.
- McGovern, C.G.: Comparing population pyramids: two techniques and a test on Alta California data from 1790 and the 1830s (=Disc. pap. ser. Dep. of geogr. Syracuse univ. 53), Syracuse, 1978.
- Raben, J. / Marks, G. (eds.): Data bases in the humanities and social sciences, Amsterdam, 1980.
- Sachs, P.: Die Bevölkerung Nenzingens im 18. Jahrhundert. (...), in: SVGB 102 (1984), 139-152.
- Schofield, R.S.: La reconstitution des familles par ordinateur, in: Annales E.S.C. 27 (1972), 1071-1082.
- Sundin, J. / Soederlund, E. (eds.): Time, space, and man. Essays on micro-demography, Stockholm, 1979.
- Tilly, C. (ed.): Historical studies of changing fertility, Princeton, 1978.
- Vetter, M.: Aufbau betrieblicher Informationssysteme, Stuttgart, 1982.
- Wachter, K.W.: Age pyramid variances, In: *ders.*: Statistical studies, New York, 1978.
- Wachter, K.W.: Statistical studies of historical structure, New York, 1978.
- Walle, E.v.d.: Démographie historique et modèles, In: Marcilio, M.L. / Charbonneau, H.: Démographie, Rouen / Montreal, 1979, 131-151.
- Wedekind, H.: Datenbanksysteme I. Eine konstruktive Einführung in die Datenverarbeitung in Wirtschaft und Verwaltung (=BI Reihe Informat. 16), 2. Aufl., Zürich, 1981.
- Willigan, J.D. / Lynch, K.A.: Sources and methods of historical demography, New York, 1982.
- Wrigley, E.A. / Schofield, R.S.: English population history from family reconstitution: summary results 1600 - 1799, in: PS 37 (1983), 157-184.
- Wrigley, E.A. / Schofield, R.S.: The population history of England, London, 1981.





## Urbarauswertung mit Hilfe des Computers

### 1. Problemskizze

Die Beschäftigung des Verfassers mit der Problematik der Urbarauswertung mittels Computer steht im Zusammenhang mit dem an der Universität Konstanz durchgeführten Forschungsvorhaben "Regionale Transformation von Wirtschaft und Gesellschaft in der frühen Neuzeit im westlichen Bodenseeraum". In diesem Projekt wird der Transformationsprozeß der vorindustriellen Gesellschaft und Wirtschaft (als Vorbedingung der kapitalistischen Industrialisierung) am Beispiel des westlichen Bodenseeraums (Hegaus) untersucht. Arbeitsschwerpunkte des "Regionalismusprojektes" sind:

1. Demographische und sozialgeschichtliche Forschungen;
2. Wirtschaftsgeschichtliche Untersuchungen;
3. Ausarbeitung eines Regionalismusmodells.

Es wird u.a. mit statistischen Methoden und mit EDV gearbeitet.

Für wirtschafts- und sozialgeschichtliche Untersuchungen in der frühen Neuzeit können Urbare eine bedeutende Rolle spielen. Diese waren jedoch wegen ihres großen Umfangs und des Überreichtums an Detailinformationen bislang nur selten Gegenstand statistischer Auswertungen, obwohl neuzeitliche Urbare zahlenmäßig viel häufiger auftreten als mittelalterliche<sup>1</sup>. Deshalb ist zunächst nicht verständlich, warum die Urbarforschung bislang als Domäne der Mediävistik anzusehen ist.

Bei der statistischen Auswertung neuzeitlicher Urbare eröffnet die EDV Möglichkeiten, die die herkömmliche Handauswertungsmethode weit übertreffen. Die Hauptschwierigkeit liegt darin, die von der äußeren Gestalt her sehr unregelmäßige historische Quelle "Urbar" ohne Veränderung des Informationsgehaltes so umzuformen, daß gängige statistische Datenanalysesysteme (wie z.B. SPSS oder BMDP) angewendet werden können.

Die vier entscheidenden methodischen Schritte der statistischen Urbarauswertung sind:

1. Einteilung der Urbareinträge<sup>2</sup> in inhaltliche Kategorien;
2. getrenntes Aufnehmen jedes einzelnen Urbareintrages;

---

<sup>1</sup> Im Hauptstaatsarchiv Stuttgart liegt der Anteil der Lagerbücher, die vor 1500 entstanden sind, nur bei etwa 2%. Vgl. Richter, Publikationsmöglichkeiten, S. 233.

<sup>2</sup> Als Urbareintrag zählt bei dieser Vorgehensweise jede im Urbar verzeichnete Parzelle, aber auch jede Abgabe usw.

3. Aufarbeiten der so entstandenen Rohdaten per Computer durch das "Fensterprogramm";
4. statistische Auswertung durch Datenanalyzesysteme.

Die hier vorzustellende Methode wurde zunächst von den Anforderungen eines konkreten Urbars (Steuerurbar des Hegauortes Orsingen von 1758) her entwickelt und durch den Vergleich mit anderen Urbartypen (z.B. mit Herrschaftsurbaren) verallgemeinert. Deshalb können mit dieser Vorgehensweise alle Varianten neuzeitlicher Urbare statistisch ausgewertet werden. Methode und erste inhaltliche Ergebnisse wurden in Form einer Studienabschlußarbeit niedergelegt<sup>3</sup>.

Die sachliche Zielsetzung dieser Form der Urbarauswertung ist eine detaillierte Untersuchung der Besitz- und Agrarstruktur zunächst eines Ortes und danach verallgemeinert einer "Region". Inhaltliche Ziele der Urbarauswertung<sup>4</sup> liegen in folgenden Bereichen:

1. Zusammenfassender Überblick über die und Beschreibung der agrarischen Situation (Anteil an Eigengut und Lehengut; Anteil der Gärten, Wiesen, Äcker usw.) in verschiedenen Hegauorten in der Mitte des 18. Jahrhunderts.
2. Behandlung der Frage von Gleich- bzw. Ungleichverteilung des bewirtschafteten Landes und Untersuchung der Betriebsgrößenstruktur.
3. Untersuchungen zur Abgabensituation.
4. Mit Hilfe von weiteren Quellen, vor allem zur Demographie, Behandlung von Fragen zur Sozialstruktur.
5. Untersuchung der Frage, welche äußeren Faktoren (z.B. Herrschaftsstruktur, Erbsitten, topographische Gegebenheiten, Klima usw.) Boden-

---

<sup>3</sup> Eine Kurzfassung dieser Arbeit ist zu finden unter Schrenk, C.: Methoden der Auswertung.

<sup>4</sup> Diese Überlegungen stehen teilweise in der Tradition v. Inama-Sternegg: Urbaren und ders.: Abhandlungen, sowie Lamprecht: Wirtschaftsleben, Bd. 2, S. 145-167, die sich jedoch beide nur mit mittelalterlichen Urbaren beschäftigen. Da der Schwerpunkt der oben erwähnten Studienabschlußarbeit in methodischen Fragestellungen lag, konnten für das Hegaudorf Orsingen zunächst nur die ersten drei der im folgenden beschriebenen sechs Punkte untersucht werden. Trotzdem liefert die quantifizierende Auswertung des Orsinger Urbars eine tragfähige Grundlage, auf der weiterführende Untersuchungen – insbesondere zur ländlichen Wirtschafts- und Sozialstruktur der vorindustriellen Zeit – aufbauen können. Diese und weitere Punkte werden zur Zeit für Orsingen und andere Hegauorte in einer weiterführenden Arbeit (Dissertation) mit dem Arbeitstitel "Untersuchungen zur Agrarstruktur des Hegaus anhand frühneuzeitlicher Urbare" untersucht.

nutzungsverhältnisse, Besitzrechtsstruktur und Betriebsgrößenstruktur eines Ortes bestimmen.

6. Versuch, die Ergebnisse von einzelnen Hegauorten auf die gesamte "Region Hegau" zu verallgemeinern und mit anderen Regionen zu vergleichen.

Es handelt sich um eine regional angelegte Querschnittsuntersuchung, nicht um einen zeitlichen Längsschnitt. Nicht gedacht ist an siedlungsgenetische<sup>5</sup> oder sprachgeschichtliche Forschungen<sup>6</sup> oder an die Untersuchung rechtsgeschichtlicher Fragestellungen<sup>7</sup>.

Die Gesichtspunkte einer solchen agrargeschichtlichen Untersuchung hängen somit auf mannigfache Weise mit verschiedenen anderen Faktoren zusammen, die das Leben der Menschen im hier untersuchten 18. Jahrhundert beeinflussen. So steht die politische Struktur, charakterisiert durch die Begriffe Verfassung und Herrschaft – beispielsweise über das geltende Erbrecht – in engem Zusammenhang mit der Betriebsgrößenstruktur eines Ortes. Aus der Kenntnis der Betriebsgrößen lassen sich in einer agrarischen Gesellschaft Rückschlüsse auf die gesellschaftlich-soziale Schichtung ziehen. Betriebsgrößenstruktur und agrarische Nutzung der landwirtschaftlichen Flächen stehen ihrerseits wiederum in engem Zusammenhang mit der demographischen Struktur des untersuchten Ortes: je fruchtbarer die landwirtschaftliche Nutzfläche, bzw. je höher der Hektarertrag, desto mehr Menschen können in einer Gesellschaft mit nach wie vor sehr engem Nahrungsspielraum ernährt werden.

Darüberhinaus hängt die Agrarstruktur eines Ortes auch vom Klima und den Reliefverhältnissen ab; jede Fruchtart stellt Mindestanforderungen an Witterung und Bodenbeschaffenheit. Ein auf die agrarische Struktur eines Orts einflußreicher Faktor ist auch dessen Marktverflechtung: einerseits stimulieren Märkte zum Erzeugen der dort nachgefragten Produkte, und andererseits benötigen landwirtschaftliche (und gewerbliche) Güter i.a. einen Absatzmarkt, der über die Grenzen des eigenen Dorfes hinausreicht. Mit Hilfe von Wochenmarktsprotokollen – ergänzt durch Zehntrechnungen und Gredbücher – läßt sich neben der Marktverflechtung eines Ortes auch der dort herrschende Intensitätsgrad der Bodennutzung ermitteln.

Alle diese Faktoren wirken zusammen, wenn nach der Zahl der Arbeitskräfte gefragt wird, die für sich entwickelndes Gewerbe oder für aufkeimende Industrie erforderlich sind, bzw. zur Verfügung stehen. Je mehr

---

<sup>5</sup> Wie z.B. bei Ott, *Agrarverfassung oder Dollinger, Bauernstand*.

<sup>6</sup> Wie z.B. bei Kleiber, *Urbare und ders., Deutsche Sprache*

<sup>7</sup> Wie z.B. bei Richter, *Kompetenzstreitigkeiten*.

Personen ihre Subsistenz nicht durch landwirtschaftliche Tätigkeit sichern können, desto schneller und intensiver kann der (Proto-) Industrialisierungsprozeß oder die Kommerzialisierung der Landwirtschaft verlaufen.

Die Kenntnis der agrarischen Verhältnisse ist also ein wichtiger Schritt auf dem Weg zur Erforschung der alltäglichen Lebensumstände des "kleinen Mannes" im 18. Jahrhundert als Voraussetzung für die Industrialisierung, durch welche unsere heutige Gesellschaft geprägt ist.

## **2. Allgemeine Gedanken zur Urbarauswertung mittels Computer**

### **2.1. Warum Urbarauswertung mit dem Computer?**

Für die Geschichtswissenschaft bedeutet der Einsatz des Computers, daß große Datenmengen, wie sie beispielsweise bei quantifizierenden Methoden anfallen, leichter bewältigt werden können als bei einer Auswertung mittels Zettelkasten. Es ist hier jedoch nicht der Ort, zu Möglichkeiten und Grenzen des Computers und der quantifizierenden Methode in der Geschichtswissenschaft Stellung zu nehmen<sup>8</sup>.

Ob eine quantifizierende Urbarauswertung überhaupt sinnvoll ist, hängt vom Informationsgehalt der Quelle bzw. von der Differenziertheit der Angaben ab. So ist es z.B. kaum angebracht, frühe mittelalterliche Urbare in einer Weise statistisch auszuwerten, wie sie hier für neuzeitliche Urbare beschrieben werden soll. Denn die quantifizierbaren Informationen in frühen mittelalterlichen Urbaren sind i.a. so ungenau, daß sich eine auf statistischen Methoden fußende Auswertung, die über das grobe Ermitteln von Summen und das Auszählen von Fällen (cases) hinausgeht, von selbst verbietet.

Neuzeitliche Urbare weisen oft einen sehr großen Umfang auf<sup>9</sup>. Die inzwischen vom Verfasser bearbeiteten Urbare belaufen sich auf bis zu 8000 Einträge. Es ist in der Praxis schwer durchführbar, diese 8000 Einträge, aufgenommen auf einzelne Karteikarten, auch nur nach einem bestimmten Kriterium zu sortieren, geschweige denn Summen oder gar Varianzen zu errechnen. Wenn man weiter bedenkt, daß es mit dem Sortieren nach einem

---

<sup>8</sup> Vgl. dazu u.a. den Bericht des oben genannten "Regionalismusprojekts": Göttmann, F./ Rabe, H./ Sieglerschmidt, J.: Regionale Transformation, hier S. 124 f.

<sup>9</sup> Urbare aus dem 17. und 18. Jahrhundert können je nach der Größe des betreffenden Ortes 1000 bis 2000 Seiten oder mehr umfassen. Mittelalterliche Urbare sind dagegen i.a. sehr viel kleiner und können deshalb auch "von Hand" ausgewertet werden.

einziges Kriterium bei weitem noch nicht getan ist, kann man ermessen, daß es eine Lebensaufgabe wäre, auch nur ein einziges neuzeitliches Urbar von Hand unter all den Gesichtspunkten auszuwerten, die Urbare als wichtige Quelle der Sozial- und Wirtschaftsgeschichte erscheinen lassen.

Der hohe Zeitaufwand für die Entwicklung der Auswertungsmethode per Computer ist gerechtfertigt, weil die Methode sowohl auf Personalprinzipurbare<sup>10</sup> als auch auf Realprinzipurbare<sup>11</sup>, aber auch auf Herrschaftsurbare<sup>12</sup>, Gesamturbare (oder Gemeindeurbare)<sup>13</sup>, Steuerurbare<sup>14</sup>, Zehnturbare<sup>15</sup> usw. angewendet werden kann. Die oben skizzierten Möglichkeiten und Vorteile legen es nahe, den Computer als Hilfsmittel für eine umfassende statistische Auswertung frühneuzeitlicher Urbare einzusetzen. Man kann sogar behaupten, daß der Informationsgehalt größerer neuzeitlicher Urbare im Grunde erst mittels Computer vollständig ausgeschöpft werden

---

<sup>10</sup> Der wesentliche Ordnungsgesichtspunkt bei Personalprinzipurbaren sind die besitzenden Bewohner eines oder mehrerer Orte. Die Personen werden nacheinander alle einzeln abgehandelt. Dabei beginnt das Urbar für jeden Besitzer beispielsweise mit den Gärten, dann folgen die Rebländer, Wiesen, Äcker usw.

<sup>11</sup> Beim Realprinzip wird zunächst nach der Bodennutzung gegliedert, also z.B. nach den Gärten, Rebländern, Wiesen, Äckern usw. Innerhalb der einzelnen Bodennutzungsarten sind die jeweiligen Parzellen nach Besitzern geordnet, wobei die Reihenfolge der Besitzer innerhalb der Bodennutzungsarten im Normalfall unverändert bleibt.

<sup>12</sup> Bei Herrschaftsurbaren handelt es sich um Quellen, die alle Rechte einer Herrschaft (Ritter, Klöster usw.) oder eines Herrschaftsteiles an Land und Leuten schriftlich erfassen und abgrenzen. Mit diesem Urbartyp war das Streben der Herrschaft nach genauen und zuverlässigen Informationen über ihre Rechte (und Pflichten) verbunden. Eigenbesitz der Bauern oder Besitz anderer Herrschaften bzw. der Kirche sind in Herrschaftsurbaren nicht behandelt.

<sup>13</sup> Gesamturbare (oder Gemeindeurbare) verzeichnen sämtlichen immobilien Besitz eventuell verschiedener Herrschaften sowie den Eigenbesitz der Bevölkerung in einem Ort und nicht nur den Anteil, der einer bestimmten Herrschaft gehört. Dadurch wird deutlich, wer wieviel Besitz in einem Ort hat und welche Bewohner welchen Herrschaften gegenüber steuerbar sind.

<sup>14</sup> Steuerurbare wurden in der frühen Neuzeit angelegt, um die Reichssteuern umlegen zu können. Ähnlich wie Gesamturbare (oder Gemeindeurbare) beziehen sie sich auf einen bestimmten Ort, verzeichnen aber nur denjenigen Besitz, der seit jeher Steuern trägt. Eigenbesitz des Adels und der Kirche bleibt unberücksichtigt.

<sup>15</sup> Zehnturbare verzeichnen alle zehnbaren Parzellen (meist innerhalb eines Ortes).

kann. Da die Computertechnik jedoch erst seit einigen Jahren die technischen Möglichkeiten für eine maschinelle statistische Urbarauswertung bietet, ist es verständlich, daß die Urbarforschung bislang eine Domäne der Mediävistik war.

## 2.2. Urbare und Computer

Obwohl die technischen Möglichkeiten der maschinellen Auswertung die der traditionellen Handauswertung bei weitem übertreffen, wurde eine Urbarauswertung mittels Computer bis heute noch kaum ernsthaft versucht. In der Literatur finden sich bislang lediglich zwei Projekte, die sich konkret mit diesem Thema beschäftigen: *R. Kropf* (1969) und *U. Portmann* (1977) <sup>16</sup>.

*Rudolf Kropf* zeigte 1969 in einem kurzen Aufsatz <sup>17</sup> die Probleme auf, denen sich ein Historiker bei dem Versuch gegenübersteht, Urbare mittels Computer auszuwerten. *Kropf* hat seine Vorstellungen jedoch nicht in die Tat umgesetzt.

Im Gegensatz dazu gelang es *Urs Portmann* <sup>18</sup> ein knappes Jahrzehnt später, eine Urbarauswertung mittels Computer durchzuführen. *Portmanns* Vorgehensweise ist jedoch auf mittelalterliche Urbare zugeschnitten und läßt sich nicht auf die sich davon stark unterscheidenden neuzeitlichen Urbare übertragen. Einschränkend sagt *Portmann* selbst, daß bei seiner Arbeit der "geschichtliche Aspekt" in den Hintergrund trete <sup>19</sup>, zugunsten der Entwicklung und Darstellung seiner Methode. *Portmann* hält es deshalb bei seinen Urbarauswertungen für vertretbar, nur eine Stichprobe <sup>20</sup> auszuwählen, die

---

<sup>16</sup> 1. *Rudolf Kropf* veröffentlichte 1969 den Aufsatz "Zur Auswertung von Urbaren mittels Computer". 2. *Urs Portmann* legte 1977 eine Lizentiatenarbeit vor mit dem Titel "Elektronische Datenverarbeitung und mittelalterliche Urbare. Auswertungen mit Urbaren des Stiftes Beromünster aus dem 14. Jahrhundert." Auf beide Arbeiten ist hier einzugehen. Am Rande soll aber auch noch der wenige Zeilen lange Hinweis aus der Zeitschrift "Annales" erwähnt werden, in welchem *Robert Fossier* (Informatique, S. 1135) auf die Idee und die damit verbundenen technischen Schwierigkeiten hinweist, Urbare maschinell (also mittels Computer) auszuwerten. Nach *Fossiers* Meinung erhält der Inhalt der Urbare dann eine besonders große Bedeutung, wenn man viele Exemplare dieser Quellenart auswertet.

<sup>17</sup> *Kropf*, Urbarauswertung.

<sup>18</sup> *Portmann*, EDV und Urbare.

<sup>19</sup> *Portmann*, ebda., S. 2.

<sup>20</sup> Es ist m.E. nicht sinnvoll, ein Urbar nur stichprobenartig auszuwerten, denn nur repräsentative Stichproben lassen Rückschlüsse auf die Grundgesamtheit zu.

nicht repräsentativ ist<sup>21</sup>. Damit gelten *Portmanns* inhaltliche Aussagen nur für die ausgewählte Stichprobe, nicht jedoch für seinen Gesamtdatenbestand<sup>22</sup>.

### 3. Methode der Urbarauswertung

Die Computermethode für die Urbarauswertung läßt sich in drei Hauptabschnitte gliedern. Zuerst muß das Urbar "computergerecht" aufgearbeitet werden, im zweiten Teil folgen die statistischen Auswertungen. Als drittes können die ermittelten Ergebnisse zur besseren Interpretierbarkeit und Anschaulichkeit in Graphiken und Tabellen dargestellt werden. Der erste Schritt ist der Kernpunkt der hier vorgestellten Urbarauswertungsmethode. Deshalb muß das "computergerechte" Aufarbeiten von Urbaren detailliert beschrieben werden.

#### 3.1. "Computergerechtes" Aufarbeiten eines Urbars

Es ist in diesem Abschnitt darzustellen, wie man die Informationen der Urbare so systematisieren kann, daß ein Abspeichern auf Datenträger und eine statistische Auswertung möglich wird. Dabei treten folgende Hauptprobleme auf:

1. Ein Urbar enthält inhaltlich verschiedenartige Informationen<sup>23</sup>.
2. Die Anzahl der Einträge (Parzellen etc.) variiert von Person zu Person und von Bodennutzung zu Bodennutzung sehr stark.

Das Urbar hat also keine einheitliche äußere Form, in der es direkt in den Rechner übertragen werden könnte. Um die vorhandenen statistischen Programmpakete<sup>24</sup> anwenden zu können, ist aber eine einheitliche äußere (Rechteck-) Form (also eine Matrix) der zu erstellenden Datei Grundvoraussetzung. Die gängigen Datenanalysesysteme sind beispielsweise auf die

---

Da aber im Falle eines Urbars die Grundgesamtheit (in Abhängigkeit von der Fragestellung) meist sehr inhomogen ist (z.B. bei der Untersuchung der Betriebsgrößenstruktur), kann eine repräsentative Stichprobe erst dann gezogen werden, wenn der Informationsgehalt der Quelle im Detail bekannt ist. Dieses Kenntnis ist aber im allgemeinen erst das Ziel der Auswertung.

<sup>21</sup> *Portmann*, EDV und Urbare, S. 14.

<sup>22</sup> *Portmann*, ebda., S. 14.

<sup>23</sup> In einem Urbar finden sich z.B. neben der Aufzählung und Beschreibung von Parzellen auch Angaben über Abgabeverhältnisse, lebensrechtliche Angelegenheiten usw.

<sup>24</sup> Hier sind vor allem SPSS und BMDP zu nennen.



	Frage 1	Frage 2	Frage 3	Frage 4
Person 1	ja	nein	ja	weiß nicht
Person 2	ja	ja	weiß nicht	ja
Person 3	nein	nein	ja	nein
Person 4	nein	nein	nein	nein
Person 5	ja	ja	nein	ja

Schema 1

	Wiesen		Rebländer	Äcker	
Person 1	Wiese 1	Wiese 2	Rebland 1	Acker 1	Acker 2
Person 2	Wiesen		Äcker	Abgaben	
	Wiese 1		Acker 1	1 Huhn	
Person 3	Rebländer		Äcker	Brachländer	
	Rebland 1	Rebland 2	Acker 1	Brachland 1	

Schema 2

Auswertung von Fragebogen zugeschnitten. In diesem Fall liegen für jede interviewte Person Antworten auf jeweils die gleichen Fragen vor<sup>25</sup>. Daraus ergibt sich ein Rechteckschema, das aussehen könnte wie Schema 1.

Bei Urbaren ist die Situation wesentlich komplizierter. Hier ergeben sich für die Einzelpersonen inhaltlich und anzahlmäßig völlig verschiedene Verhältnisse, wie sie in Schema 2 dargestellt werden.

Es ist relativ schwierig, die unregelmäßige äußere Form des Urbars in ein Rechteckschema umzuwandeln. Es wäre nicht zweckmäßig, eine Höchstanzahl von Wiesen, Äckern, Rebländern usw. festzulegen und dann die Parzellen in einem entsprechend großen Rechteckschema anzuordnen. Abgesehen von der Schwierigkeit, diese Höchstzahlen zu ermitteln, benötigte dieses Schema sehr viel (Speicher)-Platz, denn zahlreiche Positionen wären unbesetzt<sup>26</sup>.

<sup>25</sup> Man beachte, daß beispielsweise die Verweigerung einer Antwort auch eine Antwort ist.

<sup>26</sup> Gegen diesen Lösungsvorschlag spricht auch die unten geäußerte Maxime, daß

Die äußere Form der Datenaufnahme muß insgesamt verschiedenen Bedingungen genügen:

1. Die Quelle soll (bis auf Verschlüsselungen) möglichst originalgetreu wiedergegeben werden.
2. Bei der Datenaufnahme ist ein Rechteckschema zu errichten, um die gängigen statistischen Datenanalysysteme einsetzen zu können.

Die zentralen Verfahrensschritte, die dazu dienen, die oben genannten Bedingungen zu erfüllen, sind:

1. Die Aufteilung der Urbareinträge in Kategorien (S. 3.1.1.);
2. Abspeichern jedes Urbareintrages als separaten Datensatz (S. 3.1.2.);
3. Aufbereitung der Daten durch das "Fensterprogramm" (S. 3.1.3.).

### 3.1.1. Kategorien

Die Einträge der bisher vom Verfasser behandelten neuzeitlichen Urbare lassen sich im allgemeinen inhaltlich in bis zu fünf verschiedene *Kategorien*<sup>27</sup> einteilen: außer Einleitung und Index finden sich dort 1. Parzellen, 2. Abgaben, 3. Zwischen- und Endsummen, 4. die "Anlage" (ertragsorientierte Steuer) und 5. Gemeinschaftsparzellen (Parzellen, die im Besitz verschiedener Personen sind, wobei der Anteil der Einzelpersonen nicht geklärt ist)<sup>28</sup>. Eine derartige Vielfalt erschwert die Computerbearbeitung der Urbare. Deshalb werden alle Urbareinträge in inhaltlich verschiedene Kategorien aufgeteilt. Das Urbar gliedert sich auf diese Weise (vorläufig) in verschiedene Teile auf, die getrennt (und später in allen Kombinationen) ausgewertet werden können. Die notwendige Zuordnung der Daten erfolgt über Parzellennummern<sup>29</sup>.

Der erste Vorteil der Einführung von Kategorien liegt darin, daß Datensätze, die z.B. nur Parzellen enthalten, wesentlich einfacher (etwa nach

---

die Quelle im ersten Verarbeitungsabschnitt möglichst originalgetreu (und ohne Hinzufügung vieler Parzellen der Größe und des Wertes Null) aufgenommen werden sollte.

<sup>27</sup> Dabei gilt:

1. Nicht in jedem Urbar tauchen alle Kategorien auf.
2. In weiteren Urbaren können sich noch zusätzliche Kategorien finden (z.B. Leibeigenenlisten). Diese werden dann analog behandelt.

<sup>28</sup> In dem unten als Beispiel behandelten Gottmadinger Urbar von 1761 treten nur die Kategorien 1 bis 4 auf.

<sup>29</sup> Diese Parzellennummer ist in neuzeitlichen Urbaren meist schon vorhanden.

Bodennutzung) ausgewertet werden können, als die Gesamtmenge aller Urbareinträge<sup>30</sup>. Alle Kategorien werden entsprechend ihres Informationsgehaltes zunächst getrennt voneinander ausgewertet (d.h. Parzellen für sich, Abgaben für sich usw.). Die Kategorien können durch geschickten Aufbau aber auch in jeder beliebigen Kombination (d.h. z.B. Abgaben bezogen auf Parzellen) bearbeitet werden.

Der zweite Vorteil liegt im Bereich der Datenaufnahme. Wollte man z.B. Parzellen und Abgaben zusammen aufnehmen, müßte man bei jeder Parzelle die Möglichkeit offenhalten, daß Abgaben darauf haften<sup>31</sup>. Das Datenaufnahmesystem würde auf diese Weise unnötig kompliziert und aufgebläht<sup>32</sup>.

### 3.1.2. Separate Datensätze

Der zweite Kerngedanke betrifft die Datenaufnahme innerhalb der einzelnen Kategorien. Da nicht von vornherein bekannt ist, ob eine Person zwei oder 20 "gleichartige" Parzellen (z.B. Eigenwiesen) besitzt, hat es sich als am zweckmäßigsten herausgestellt, jeden Urbareintrag als einen separaten Datensatz abzuspeichern<sup>33</sup>.

Zweckmäßig ist diese Vorgehensweise hauptsächlich auch deshalb, weil durch die Aufnahme separater Datensätze der Unterschied zwischen Personal- und Realprinzipiurbaren verarbeitungstechnisch keine Rolle mehr spielt. Mit einem einfachen Sortiervorgang<sup>34</sup> können mittels Computer Real- und Personalprinzipiurbare wechselseitig ineinander übergeführt werden, wenn alle Datensätze einzeln verschlüsselt sind<sup>35</sup>.

---

<sup>30</sup> So sind z.B. Informationen über die Abgaben für die Frage nach der Nutzung zunächst irrelevant oder gar störend.

<sup>31</sup> Auch hier ergäbe sich u.a. wieder das Problem einer Höchstanzahl von Abgaben pro Parzelle.

<sup>32</sup> Vgl. dazu unten 4.3

<sup>33</sup> Eine weitere Möglichkeit bestünde z.B. darin, mehrere Parzellen zu einem Datensatz zusammenzufassen, was jedoch schon aus Gründen der Übersichtlichkeit nicht sinnvoll ist.

<sup>34</sup> Das Sortieren gehört zu den gängigen Kommandos und ist i.a. in den Betriebssystemen verfügbar. Die ausführlichste Übersicht über Sortierv Verfahren bietet Knuth, D.E.: The Art of Computer Programming, Bd. 3: Sorting and Searching.

<sup>35</sup> Die Zuordnung etwa von Abgaben und Parzellen geschieht mittels einer Parzellennummer, die, wie bereits erwähnt, in neuzeitlichen Urbaren meist schon in der Quelle vorhanden ist. Die Nummer stimmt meist mit einem üblicherweise gleichzeitig zum Urbar angelegten Feldriß überein.

Person 1  
Wiese 1  
Wiese 2  
Acker 1  
Abgabe 1  
Abgabe 2  
Acker 2  
usw.

Person2  
analog  
usw.

### Schema 3

Zusammenfassend gelten für die äußere Form der Datenaufnahme die beiden Grundsätze, daß (1.) jeder Urbareintrag als separater Datensatz verschlüsselt wird und daß (2.) jeder Urbareintrag einer bestimmten Kategorie zugeordnet ist. Auf diese Weise entsteht ein Rechteckschema "im weiteren Sinne"<sup>36</sup>, wie in Schema 3 dargestellt.

Durch die separate Aufnahme jedes einzelnen Eintrages in der Reihenfolge des Urbars wird die Quelle weitgehend originalnah wiedergegeben.

#### 3.1.3. Das "Fensterprogramm"

Das selbstgeschriebene<sup>37</sup> zentrale Programm für die Urbarauswertung heißt "Fensterprogramm"<sup>38</sup>. Dieses Programm wurde in der Computerspra-

---

<sup>36</sup> Der Zusatz "im weiteren Sinne" soll andeuten, daß nicht (wie bei einem Fragebogen) zu jeder Person gleichviele Informationen vorliegen, sondern daß sich die Anzahl der zugehörigen Urbareinträge ändert. Ein strenges Rechteckschema wird erst mit Hilfe des "Fensterprogramms" erzielt (vgl. 3.1.3.).

<sup>37</sup> Für wertvolle Hilfe beim Schreiben des "Fensterprogramms" danke ich herzlich Frau cand. wirt. *Elke Sengteller*. - Eine für die Veröffentlichung gedachte Neufassung des Programms entwickelte Dipl. Phys. *Otto Stolz*, dem ich für seine Mühe ebenfalls herzlich danke.

<sup>38</sup> Das Fensterprogramm wurde inzwischen in verschiedenen Versionen für völlig andere historische Fragestellungen eingesetzt, z.B. für die Auswertung von Wochenmarktprotokollen, Bürgerlisten, Preisreihen usw.

Person	Besitzrechtsform	Bodennutzung	Größe (in Hektar)	Wert (in Gulden)
Britsch	Eigengut	Wiese	1	50
Britsch	Eigengut	Wiese	3	100
Britsch	Lehengut	Wiese	2	80
Britsch	Eigengut	Acker	3	100
Britsch	Eigengut	Acker	2	70
Britsch	Eigengut	Acker	4	150
Fahr	Lehengut	Wiese	1	40
Fahr	Lehengut	Wiese	2	80
Fahr	Lehengut	Rebland	1	200
Fahr	Lehengut	Rebland	0.5	100

Schema 4

che Pascal<sup>39</sup> geschrieben. Es bringt die Urbarinformationen in ein "strenges Rechteckschema" und leistet eine erste statistische Auswertung des Urbars. Da ohne das Fensterprogramm oder etwas Entsprechendes die Urbarauswertung in dieser Weise nicht möglich gewesen wäre, sei seine Funktionsweise anhand eines Beispiels erläutert<sup>40</sup>.

Es ist ein wichtiger Auswertungsschritt zu berechnen, wieviel Land jeder im Urbar aufgeführten Personen zur Verfügung steht. Dabei sollen die Angaben nach Besitzrechtsform (Eigengut, Lehen usw.) und nach Bodennutzung (Wiese, Rebland, Acker usw.) aufgegliedert sein. Es wird gefragt nach der Anzahl der jeweiligen Parzellen und deren Gesamtgröße, Durchschnittsgröße, auftretendem Minimum und Maximum, Varianz usw. Entsprechende Angaben sollen auch über den im Urbar oft angegebenen Steuerwert der Parzellen errechnet werden.

Das fiktive Beispiel in Schema 4 soll das Gesagte verdeutlichen.

<sup>39</sup> Pascal ist m.E. für das hier zu behandelnde Problem der beste Kompromiß zwischen bequemer und übersichtlicher Programmiersprache sowie möglichst weiter Verbreitung. Pascalcompiler gibt es für fast alle Groß- und Kleinrechner.

<sup>40</sup> Es ist zu betonen, daß das hier vorgestellte "Fensterprogramm" bei weitem nicht der einzige programmiertechnische Lösungsweg ist. Als weitere Möglichkeit sei an Datenbanken und Netzwerksysteme erinnert (die jedoch nicht auf jedem Computer "laufen"). Nicht zu vergessen sind auch die im kommerziellen Bereich angebotenen "Gruppenwechselprogramme" (die jedoch zusätzlich Geld kosten).

Britsch	: 2 eigene Wiesen,	Gesamtgröße	4 Hektar
		Durchschnittsgröße	2 Hektar
		Minimum (Größe)	1 Hektar
		Maximum (Größe)	3 Hektar
		Varianz (Größe)	2.000
		Standardabweichung	1.414
		Gesamtwert	150 Gulden
		Durchschnittswert	75 Gulden
		Minimum (Wert)	50 Gulden
		Maximum (Wert)	100 Gulden
		Varianz (Größe)	1250.00
		Standardabweichung	35.36
		Durchschnittswert (gewichtet)	
		in Gulden pro Hektar	37.5
Fahr	: 1 Lehenswiese,	Angaben analog	
	: 3 eigene Äcker,	Angaben analog	
	: 2 Lehenswiesen,	Angaben analog	
	: 2 Lehenrebländer,	Angaben analog	

## Schema 5

Der Computer soll die Informationen aus Schema 5 liefern.

Dieser Aufwand wäre kaum lohnend, wenn es sich pro Person nur um zwei oder drei Parzellen handelte; in den Urbaren finden sich jedoch oft Personen mit über 50 Parzellen.

Der Computer muß die durch das Fensterprogramm errechneten Informationen zur späteren Weiterverarbeitung (z.B. per SPSS) wieder in eine Datei eintragen. Die Ergebnisse werden dabei nicht in der oben angedeuteten Ausführlichkeit aufgelistet, sondern teilweise verschlüsselt und in einer bestimmten Reihenfolge hintereinander für jede Person in eine Zeile abgelegt, so daß für verschiedene Personen die einzelnen Variablen immer spaltenweise untereinander stehen. Damit wird ein "strenges Rechteckschema" erreicht, und die Daten können mit den gängigen Datenanalysesystemen weiterverarbeitet werden.

Der Name "Fensterprogramm" erklärt sich dadurch, daß (wie durch ein Fenster) zunächst nur bestimmte Teile der Datei betrachtet werden. (Im obigen Beispiel wären diese Teile der Name, die Besitzrechtsform und die

Bodennutzung.) Das "Fenster" kann aus maximal fünf<sup>41</sup> voneinander völlig unabhängigen oder sich überlappenden Teilen bestehen. Damit ist der Computer mit Hilfe des Programms in der Lage, eine bestimmte Zeichenkombination, die eventuell auf verschiedene Positionen in der Datei verteilt ist (also im obigen Beispiel Name, Besitzrechtsform, Bodennutzung), zu erkennen und abzuzählen. Das ist die Grundversion des Programms, die um einige für die Urbarauswertung wichtige statistische Parameter (wie Summe, Durchschnitt, Varianz usw.) erweitert wurde. Der Einsatz des Programms setzt eine in Spaltenform<sup>42</sup> aufgebaute und nach den entsprechenden Kriterien (Name, Besitzrechtsform, Bodennutzung) sortierte Datei voraus<sup>43</sup>.

In die Urbarauswertung fließen noch verschiedene andere Programme mit ein, die beispielsweise die Zahlenangaben im Urbar<sup>44</sup> in Dezimalwerte umrechnen. Weitere Programme dienen dem Umordnen, dem Codieren oder Decodieren von Informationen.

Vor der eigentlichen statistischen Auswertung ist es sinnvoll, die im verschlüsselten Urbar auftauchenden Größen- und Wertangaben (in Jauchert, Vierteln und Ruten bzw. Gulden, Kreuzern und Hellern) ins Dezimalsystem (in Jauchert bzw. Gulden, dargestellt als Kommazahlen) umzurechnen. Dies geschieht im vorliegenden Fall mit Hilfe dafür selbstgeschriebener Computerprogramme, ist aber beispielsweise auch mit SPSS realisierbar.

#### 3.1.4. Verschlüsselungen

Die Informationen des Urbars werden nicht im Klartext, sondern verschlüsselt in den Rechner eingegeben. Die wichtigsten Vorteile, die eine Verschlüsselung bieten kann, sollen hier aufgezählt werden.

- Durch formatierte Codierungen kann auf platzsparende Weise ein Spaltenaufbau erreicht werden, der für den Einsatz des Fensterprogramms

---

<sup>41</sup> Durch eine einfache Änderung im Programm kann diese Maximalzahl geändert werden.

<sup>42</sup> Wenn man eine Urbarauswertung etwa mit Hilfe der Tübinger Textverarbeitungsprogramme (TUSTEP) durchführt, können die Eingabedaten unformatiert sein, sie müssen jedoch mit Erkennungsmarken versehen werden.

<sup>43</sup> Allgemeine Literatur über Sortierverfahren wurde schon in Abschnitt 3.1.2. (Anm. 33) genannt. Das Sortieren im Spezialfall "Pascal" wird behandelt bei Erbs, H.-E./Stolz, O.: Einführung in die Programmierung mit Pascal. 2. Auflage, 1984, hier: Kap. 11.1 und 11.2., S. 159-172.

<sup>44</sup> In den Hegauer Urbaren sind die Grundstücksgrößen in Jauchert, Vierteln und Ruten, die Parzellenwerte in Gulden, Kreuzern und Hellern angegeben.

notwendig ist<sup>45</sup>;

- es ergibt sich eine Vereinheitlichung bei Schreibvarianten;
- die Informationen werden durch systematische Verschlüsselungen strukturiert;
- der Computereinsatz ist bei Verwendung von Ziffern problemloser als bei Buchstaben<sup>46</sup>.

Die Verschlüsselungen müssen verschiedenen Bedingungen genügen:

- Die Rekonstruierbarkeit der Quelle muß gegeben sein;
- der Codeplan ist flexibel zu halten (d.h. der Codeplan muß noch während des Quellenabschreibens abgeändert bzw. erweitert werden können);
- der Codeplan sollte prinzipiell auf verschiedene Varianten desselben Quellentyps (z.B. sowohl auf Realprinzip- als auch auf Personalprinzipurbare) angewendet werden können.

Zu den Nachteilen des Codierens zählt insbesondere, daß kleinere Fehler nur schwer zu erkennen sind.

Bei der Verschlüsselung gelten grundsätzlich folgende Konventionen:

1. Alle aufeinanderfolgenden Codierungen sind durch mindestens ein Leerzeichen zu trennen<sup>47</sup>.
2. Gleichartige Informationen, die in verschiedenen Kategorien auftreten, werden gleich verschlüsselt<sup>48</sup>.
3. Die Codierungen werden in Spaltenform angelegt<sup>49</sup>.

---

<sup>45</sup> Der Spaltenaufbau kann jedoch auch ohne Verschlüsselungen erreicht werden.

<sup>46</sup> Das bedeutet aber nicht, daß nur Codierungen verarbeitet werden können, die ausschließlich Ziffern verwenden. Man kann auch mit Buchstaben oder ganzen Worten arbeiten, nur ist das programmtechnisch oftmals aufwendiger.

<sup>47</sup> Vorteile: Wenn die codierte Quelle durch Leerzeichen strukturiert ist, so ist sie wesentlich leichter zu lesen als direkt aneinandergehängte Verschlüsselungen, und sie kann computertechnisch einfacher zu bearbeiten sein. Möglicher Nachteil: Man benötigt mehr Platz.

<sup>48</sup> Hierfür ein Beispiel: Die Bodennutzung ist eine Information, die bei allen fünf Kategorien auftritt. Somit wird in allen fünf Kategorien für dieselbe Bodennutzungsart derselbe Schlüssel verwendet, wenn auch teilweise auf verschiedenen Spalten.

<sup>49</sup> Dem spaltenförmigen Aufbau des codierten Urbars wurde aus programmtechnischen Gründen (vgl. 3.1.3.) der Vorzug gegeben vor einem System, das auf den Spaltenaufbau verzichtet, dafür aber auf das Setzen von Erkennungsmarken zur Identifikation der Variablen angewiesen ist. Dadurch, daß beim Spaltenaufbau jede Variable auf genau definierten Spalten verschlüsselt wird, kann die Variable durch ihre Position identifiziert werden. Zwar können die beiden Codierungssy-



Vor dem Erstellen eines Codeplans ist zu entscheiden, welche Teile der Quelle aufgenommen werden. Außerdem soll das Format der Schlüssel und deren genaue Position festgelegt werden<sup>50</sup>. Dabei dürfen keine relevanten Informationen verlorengehen, und spätere Auswertungsideen müssen auch nachträglich realisiert werden können.

### 3.1.5. Statistische Auswertung

Das "computergerechte" Aufarbeiten eines Urbars ist nur die notwendige Vorarbeit für die statistische Auswertung der Detailinformationen jener Quelle, die bis zu einem gewissen Grad schon durch das Fensterprogramm erzielt wird. Für die Berechnung weitergehender Statistiken, wie z.B. Korrelationen, müssen die gängigen statistischen Datenanalysesysteme herangezogen werden<sup>51</sup>. Diese statistische Auswertung ist aber ihrerseits wieder nur eine Vorarbeit für den eigentlichen Zweck der Urbarauswertung, der Beantwortung bestimmter historischer Fragen.

Es ist jedoch hier nicht der Ort, auf die Problematik der Anwendung statistischer Arbeitstechniken in der Geschichtswissenschaft einzugehen<sup>52</sup>

## 4. Konkrete Vorgehensweise

Der folgende Abschnitt zeigt beispielhaft anhand des Gottmadinger Steuerbuchs von 1761, wie das Codieren und Auswerten konkret vor sich gehen kann.

### 4.1. Der Aufbau der ausgewerteten Urbare

Bei den bislang ausgewerteten Urbaren handelt es sich hauptsächlich um Steuer- und Zehnturbare, die nach dem Personalprinzip aufgebaut sind<sup>53</sup>.

steme wechselseitig ineinander umgeformt werden, doch ist es sinnvoll, sofort einen Spaltenaufbau anzustreben, wenn dieser in der Auswertungsphase gebraucht wird.

<sup>50</sup> Eine Änderung von Format und Position der Schlüssel ist technisch jedoch ebenso jederzeit möglich, wie das Einfügen weiterer zu verschlüsselnder Variabler.

<sup>51</sup> Aus Platzgründen muß an dieser Stelle auf weiterführende Betrachtungen verzichtet werden. Als Literaturangabe – etwa für SPSS – sei genannt: *Beutel, M./Schubö, W.: SPSS 9.*

<sup>52</sup> Als Literaturangabe sei aus der Fülle der Möglichkeiten herausgegriffen: *Stenzel, I.: Fehler und Fallen.*

<sup>53</sup> Da jeder Urbareintrag als separater Datensatz verschlüsselt wird, spielt jedoch, wie in Abschnitt 3.1.2. gezeigt, der formale Aufbau des Urbars methodisch gesehen keine Rolle. Ebensowenig ändert sich die Auswertungstechnik bei Steuer-, Zehnt-, Gesamt- oder Herrschaftsurbaren.

Bei der Codierung wird davon ausgegangen, daß in dem betreffenden Urbar für jede *Parzelle* folgende Informationen vorhanden sind:

- Informationen über die Besitzrechtsform (Eigen- bzw. Lehensbesitz) und gegebenenfalls über den Lehensherren;
- Informationen über den Besitzer (Name usw.);
- Informationen über die Größe der Parzelle;  
weiter *können* vorhanden sein:
- Informationen über die Ortslage des Parzelle;
- Informationen über die steuerliche Klassifizierung und den Steuerwert<sup>54</sup> der Parzelle usw.

Bei den *Abgaben* werden z.B. folgende Informationen erwartet:

- Ware und Menge der Abgabe;
- Empfänger der Abgabe;
- Abgabepflichtiger.  
Weiter *können* vorhanden sein:
- Geldwert der Abgabe;
- Nutzung der belasteten Parzelle usw.

#### 4.2. Der Codeplan für das Gottmadinger Urbar von 1761

Anhand des Gottmadinger Steuerurbars von 1761 soll nun beispielhaft gezeigt werden, wie ein neuzeitliches Urbar computergerecht aufgearbeitet werden kann.

Um den Aussagewert der Urbare nicht vorzeitig einzuengen, wurden alle Detailinformationen aufgenommen, auf die viel Platz beanspruchende genaue verbale Lagebeschreibung<sup>55</sup> der Parzellen in der Gemarkung und die Anstößer wurde jedoch verzichtet<sup>56</sup>. Diese Informationen sind zwar für siedlungsgeschichtliche Fragestellungen interessant, aus ihnen sind aber keine Antworten auf die in Abschnitt 1 angesprochenen Fragestellungen zu erwarten<sup>57</sup>. Außerdem liegt im Falle der Urbare des 18. Jahrhunderts meist ein

---

<sup>54</sup> Der Steuerwert errechnet sich aus der Größe der Parzelle und der Klassifizierung.

<sup>55</sup> Beispielsweise heißt es im Gottmadinger Urbar von 1761 auf S. 7 im 3. Eintrag: „... liegt gegen Aufgang an Johannes Spéri, Niedergang an Joachim Kessinger eigen Acker, Mitnacht die Singer Landstraß...“.

<sup>56</sup> Vgl. Ende des ersten Abschnitts.

<sup>57</sup> Diese Vorgehensweise ist das Ergebnis einer Abwägung einerseits zwischen der wünschbaren Vollständigkeit bei der Erfassung einer Quelle für den Fall einer späteren Einspeisung der Rohdaten in eine Datenbank und andererseits der

parallel dazu angefertigter Gemeindeatlas (Feldriß) vor, aus dem sich mit Hilfe der Parzellennummern die genaue Lage aller Parzellen ablesen läßt. Das Codierungssystem kann aber in der Weise erweitert werden, daß die näheren Angaben zur Lage einer Parzelle noch nachträglich aufgenommen werden können.

Der Codeplan ist aus vier Teilen zusammengesetzt:

1. Allgemeine Angaben (z.B. Verschlüsselung der Quelle, Seitenangabe, laufende Nummer des Eintrags pro Seite);
2. Angaben zur Person des Parzellenbesitzers bzw. des Abgabepflichtigen (Name, Herkunftsort, Beruf);
3. Kategorie des Urbareintrags (Parzelle, Abgabe ...);
4. Spezialangaben (je nach Kategorie des Urbareintrages).

Die vier inhaltlichen Kategorien des Gottmadinger Urbars von 1761 spiegeln sich im Aufbau des Codeplanes wider. Während der 1., 2. und 3. Teil für alle Kategorien gleich ist, variiert der 4. Teil stark. Die Codierung der einzelnen Kategorien ist in den folgenden Abschnitten beschrieben.

#### **4.2.1. 1. Teil: Allgemeine Angaben**

Jeder Datensatz ist mit der Information zu beginnen, um welche Quelle es sich handelt. Hierfür sind die Spalten 1 und 2 vorgesehen. Auf den Spalten 4 bis 7 jedes Datensatzes wird die Seite im Urbar, auf den Spalten 9 bis 10 die Position des Eintrags auf der entsprechenden Seite abgelegt. Damit kann die Quelle und die genaue Position jedes Urbareintrags in ihr rekonstruiert werden, was u.a. für Korrekturzwecke unerlässlich ist.

#### **4.2.2. 2. Teil: Angaben zur Person**

Zu allen im Urbar auftretenden Personen werden vier Angaben aufgenommen. Auf den Spalten 12 bis 15 ist der Nachname, auf den Spalten 17 bis 19 der Vorname zu verschlüsseln. Um zu gewährleisten, daß zwei verschiedene Personen, die den gleichen Vor- und Nachnamen tragen, unterschieden werden können, wird auf den Spalten 21 bis 23 noch eine Zusatznummer geführt, mit deren Hilfe eventuell vorhandene Berufsangaben oder Zusätze wie "Wittib", "ledig" usw. verschlüsselt werden. Als weitere Angabe folgt der Herkunftsort der Person (Spalten 25 bis 27). Die Namens-, Berufs- und

---

technischen und finanziellen Machbarkeit (manpower) unter dem Aspekt, welche Informationen für die eigenen Fragestellungen relevant sind.

Ortsschlüssel sind fortlaufend gewählt und werden im erwähnten Regionalismusprojekt einheitlich verwendet<sup>58</sup>.

#### **4.2.3. 3.Teil: Dateinummern**

Bei jedem Urbareintrag wird mitverschlüsselt, zu welcher der inhaltlichen Kategorien er gehört. Es wäre möglich, aber zu umständlich, die Kategorie aus dem Kontext zu erschließen. Zur Verschlüsselung wurden die Kategorien durchnummeriert<sup>59</sup> und der Code auf den Spalten 29 und 30 abgelegt<sup>60</sup>. Durch einen einfachen Sortiervorgang können so die verschlüsselten Urbareinträge nach den verschiedenen Kategorien getrennt werden, was für die Bearbeitung von Bedeutung ist.

#### **4.2.4. 4.Teil: Spezialangaben**

Bis zur 31. Spalte ist der Codeplan für alle Kategorien gleich gestaltet. Der weitere Aufbau variiert jedoch stark und soll für die 1. Kategorie ausführlich (exemplarisch), für die anderen Kategorien kurz beschrieben werden.

##### **a) 1. Kategorie: Parzellen**

Für jede einzelne im Urbar aufgeführte Parzelle werden folgende Informationen aufgenommen:

1. Die Besitzrechtsform (Spalten 32 bis 34).
2. Die im Urbar verzeichnete Parzellennummer (Spalten 36 bis 39).
3. Zwei Spalten für Besonderheiten im Zusammenhang mit der Parzellennummer (Spalten 41 und 42).
4. Der Name des Gewanns, in welchem die Parzelle liegt (Spalten 44 bis 46).
5. Die rechtliche Bodennutzung (Spalten 48 und 49).
6. Die tatsächliche Bodennutzung (Spalten 51 bis 53).
7. Eine Sondernummer zur vereinfachten Steuerung der Computerverarbeitung (Spalten 55 bis 57).

---

<sup>58</sup> Damit können Ergebnisse von z.B. demographischen Forschungen und von Urbarauswertungen miteinander gekoppelt werden.

<sup>59</sup> "01" = Parzellen, "02" = Abgaben, "03" = Summen usw.

<sup>60</sup> Da 2 Spalten reserviert sind, besteht die Möglichkeit, noch wesentlich mehr (eventuell in anderen Urbaren auftretende Kategorien wie Leibeigenenlisten, Abgabeverzeichnisse von Fastnachtshennen usw.) aufzunehmen.

8. Die Parzellengröße in Jauchert (Spalten 59 bis 61), Vierteln (Spalte 63) und Ruten (Spalten 65 bis 70).
9. Die Wertklassifizierung der Parzellen in Kreuzern (Spalten 72 und 73) und Hellern (Spalten 75 bis 78).
10. Der Parzellenwert in Gulden (Spalten 80 bis 83), Kreuzern (Spalten 85 und 86) und Hellern (Spalten 88 bis 91).

Diese 10 Stichpunkte bedürfen der Erläuterung.

Zu 1.: Die Besitzrechtsform der Parzellen (Eigen-, Lehenbesitz usw.) wird durch eine dreistellige Zahl verschlüsselt<sup>61</sup>.

Zu 2.: Im Gottmadinger Urbar von 1761 ist jede Parzelle mit einer vierstelligen Nummer versehen. Diese Parzellennummer wird unverändert übernommen. Sie kann einen direkten Vergleich mit zeitlich anders gelagerten Urbaren oder mit anderen Quellen ermöglichen. Außerdem dient dies zur Identifikation der Parzellen innerhalb des Urbars und der Zuordnung von Parzellen und Abgaben (die getrennten Kategorien angehören)<sup>62</sup>.

Zu 3.: Verschiedentlich sind die Parzellennummern mit Zusätzen versehen, wie z.B. "1/2". Das ist hauptsächlich dann der Fall, wenn ein und dieselbe Parzelle verschiedene Besitzer hat und/oder verschieden genutzt wird; wenn es sich also um einen ehemals zusammengehörigen Besitzkomplex handelt. Ein entsprechender Zusatz kann aber auch auf einen Nachtrag oder Einschub hindeuten. Ein eventueller Zusatz wird bei der Urbarverschlüsselung mitcodiert. Beim Korrekturlesen ist diese Information nützlich<sup>63</sup>.

Zu 4.: Bei jeder im Urbar beschriebenen Parzelle ist der Name der Flur angegeben, in welcher die Parzelle liegt. Daraus läßt sich die ungefähre Lage der Parzellen ablesen (obwohl auf die genaue Lagebeschreibung mit Anstößern aus dem Urbar verzichtet wurde).

<sup>61</sup> Die 1. Stelle gibt die Besitzrechtsform an ("1" = Eigenbesitz, "2" = Lehen, "3" = Erblehen, "4" = Schupflehen, "5" = Mannlehen usw.) und die 2. Stelle den Lehensherren ("0" = kein Lehenherr (da Eigenbesitz), "1" = Kloster Reichenau, "2" = Haus Krähen, "3" = Kloster Allerheiligen in Schaffhausen usw.). Auch hier handelt es sich um eine systematische Codierung. Die dritte Stelle spezifiziert den Eigenbesitz bzw. das Lehen weiter. So bedeutet z.B. die Besitzrechtsform 531 das als Mannlehen ausgegebene "Kleine Widumgut" des Klosters Allerheiligen in Schaffhausen.

<sup>62</sup> Bei sehr großen Urbaren könnte hier eine fünfstellige Nummer auftreten, was dann im Codeplan entsprechend berücksichtigt werden müßte.

<sup>63</sup> Wenn eine Parzellennummer ohne einen entsprechenden Zusatz mehrfach auf-taucht, liegt sehr wahrscheinlich ein Übertragungsfehler vor.

Bei der Codierung der Flurnamen erwies sich eine Kombination aus systematischer und fortlaufender Verschlüsselung als sinnvoll<sup>64</sup>.

Zu 5. und 6.: Eine weitere Information über die Parzelle besteht in der Angabe der rechtlichen Bodennutzung<sup>65</sup>, die sich von der tatsächlichen Bodennutzung unterscheiden kann. So kann eine Gartenparzelle als Rebland genützt sein. Folglich sind sowohl die rechtliche als auch die tatsächliche Bodennutzung zu verschlüsseln<sup>66</sup>.

Zu 6. bis 10.: An dieser Stelle folgen eine Sondernummer und Angaben über Größe, Wertklassifizierung und Steuerwert der Parzelle<sup>67</sup>. Bis auf die

---

<sup>64</sup> Der entsprechende Schlüssel umfaßt drei Stellen. Die 1. Stelle sagt aus, ob es sich um Wald, Wiese, Ackerland usw. handelt, die 2. und 3. Stelle wurde fortlaufend vergeben. Es ist zwar nicht immer eindeutig möglich, von der ersten Stelle auf die tatsächliche Nutzung der Parzelle zurückzuschließen (da es z.B. vorkommen kann, daß eine Wiesenparzelle in einem Ackergebiet liegt). Das ist auch nicht der Sinn dieser Art der Codierung. Es geht vielmehr darum, die Vielzahl der Flurnamen zu strukturieren.

<sup>65</sup> Unter "rechtlicher Bodennutzung" soll die im Urbar durch Rubrik oder direkte Nennung angegebene bzw. rechtlich vorgegebene Form der Bodennutzung verstanden werden.

<sup>66</sup> Die Bodennutzung wird in Form einer zweistelligen Zahl fortlaufend verschlüsselt ("10" = Garten, "20" = Rebland usw.), die tatsächliche Nutzung als dreistellige Zahl. Dabei nimmt die erste Stelle bezug auf die Codierung der rechtlichen Bodennutzung, die 2. und 3. Stelle sagt Näheres über die konkrete Nutzung ("100" = Garten, "101" = Garten mit Schuppen, "201" = zusätzlich als Garten genutztes Rebland; 300 = Acker, 310 = Acker im Ösch 1, 311 = Acker und Wiese im Ösch 1, 320 = Acker im Ösch 2, 330 = Acker im Ösch 3 usw.).

<sup>67</sup> Die Rutenangabe enthält im Falle des Gottmadinger Urbars von 1761 häufig eine Bruchzahl, die beim codierenden Abschreiben in das Dezimalsystem umgewandelt wurde. Die Rutenangabe besteht somit aus drei Stellen vor und zwei Stellen nach dem Komma. Es gilt dabei: 1 Jauchert = 4 Viertel; 1 Viertel = 125 Ruten. Die Wertklassifizierung macht eine Aussage über den Steuerwert der Parzelle in Kreuzern und Hellern pro Rute. Auch hier treten bei der Hellerangabe oft Bruchzahlen auf, die ebenfalls in das Dezimalsystem umgeformt wurden. Die Hellerangabe besteht somit aus zwei Stellen vor und einer Stelle nach dem Komma. Es gilt: 1 Gulden = 60 Kreuzer; 1 Kreuzer = 8 Heller. Der im Urbar angegebene Parzellenwert läßt sich aus der Parzellengröße und der Wertklassifizierung errechnen. Auch bei den Steuerwerten treten wieder häufig Bruchzahlen auf, die in das Dezimalsystem umgewandelt wurden. Wie schon bei der Wertklassifizierung sind bei der Hellerangabe eine Stelle vor und zwei Stellen nach dem Komma vorgesehen. Unter

Umwandlung von Brüchen in Dezimalstellen werden die Größen- und Wertangaben beim Verschlüsseln nicht verändert.

Die noch vor der Parzellengröße vorgesehene zweistellige Sondernummer hat computertechnische Gründe. Sie ist zwar nicht zwingend notwendig, erleichtert aber das Programmieren ganz entscheidend. Mit ihr kann man dem Computer individuelle "Verarbeitungshinweise" geben, die immer dann wichtig sind, wenn Angaben im Urbar auftauchen, die das Codierungssystem sprengen. (Das ist allen quantifizierend arbeitenden Historikern ein wohlbekanntes Problem.) So kommt es vor, daß die Parzellengröße nicht oder nur unvollständig angegeben ist. Es kann zu schweren Fehlern in der Auswertung führen, wenn der Computer an einer bestimmten Stelle Angaben erwartet, dort aber (ausnahmsweise) keine vorhanden sind<sup>68</sup>. Problematisch sind aber nicht nur fehlende, sondern auch der Norm nicht entsprechende Angaben.

Damit ist der Codeplan für die häufigste Kategorie der Urbareinträge, die Parzellen, ausführlich beschrieben. Die Kategorien 2 bis 5 werden nur stichwortartig behandelt.

## **b) 2. Kategorie: Abgaben**

Da sich die Beschreibung der Abgaben aus anderen Teilen zusammensetzt als die der Parzellen, muß entsprechend der Codeplan (nach der 31. Spalte) anders aufgebaut sein. Folgende Informationen wurden (falls vorhanden) aufgenommen (wobei nur die Punkte 6. bis 12. der Erläuterung bedürfen, weil die Punkte 1. bis 5. schon bei der ersten Kategorie erklärt wurden<sup>69</sup>):

1. Die Parzellennummer (Spalten 34 bis 37).
2. Eine Spalte für Besonderheiten im Zusammenhang mit der Parzellennummer (Spalte 39).
3. Die Besitzrechtsform (Spalten 41 und 42).
4. Die rechtliche Bodennutzung (Spalten 44 und 45).
5. Die tatsächliche Bodennutzung (Spalten 47 bis 49).

---

"Steuerwert einer Parzelle" wird also nicht ein eventueller Grundstücks- oder Baulandpreis verstanden, sondern ein am landwirtschaftlichen Ertragswert orientierter Wert.

<sup>68</sup> Die Sondernummer "001" gibt dem Rechner die Information, daß alle Angaben vorhanden sind und dem Standard entsprechen. Sondernummer "010" bedeutet z.B., daß die Steuerwertangabe der Parzelle fehlt.

<sup>69</sup> Das entspricht dem in 3.1.4. aufgestellten Grundsatz, gleichartige Informationen in verschiedenen Kategorien gleich zu verschlüsseln.

6. Der Abgabetyp (Spalte 51 und 52)<sup>70</sup>.
7. Der Empfänger der Abgabe (Spalte 54 und 55).
8. Warenschlüssel der Abgabe (Spalten 57 und 58)<sup>71</sup>.
9. Der Zahlungsmodus (Spalten 60 und 61)<sup>72</sup>.
10. Eine Sondernummer für die Verarbeitung (Sp. 63 bis 64).
11. Der Wert der Abgabe in Gulden (Spalten 66 und 67), Kreuzern (Spalten 69 und 70) und Hellern (Spalte 72).
12. Eine Angabe (falls vorhanden) in Stück, Pfund, Malter usw. (Spalten 74 bis 80, als Kommazahl)<sup>73</sup>.

---

<sup>70</sup> Z.B. der Grundzins, Frondienste, usw.

<sup>71</sup> Der Warenschlüssel ist zweistellig. Die erste Stelle gibt einen "Gattungsbegriff" wieder ("1" = Tier, "2" = Getreide, usw.), die zweite Stelle eine genaue Angabe ("11" = Huhn, "12" = Henne, "22" = Roggen, "23" = Gerste). Falls Geld abgegeben wird, lautet der Warenschlüssel "01".

<sup>72</sup> Der Zahlungsmodus ist wichtig, weil es oft vorkommt, daß Naturalabgaben in Geld beglichen werden. Wenn der Grundzins z.B. ein Huhn beträgt, so gibt der Zahlungsmodus an, ob tatsächlich ein Huhn oder ob der Gegenwert in Geld abzugeben ist. (Zahlungsmodus = "2" bedeutet, daß, in Geld bezahlt wird, statt in Naturalien.) Das ist ein wichtiges Kriterium für die Beurteilung der Abgaben. Interessant ist dabei, ob der angesetzte Geldwert tatsächlich dem Marktwert der Naturalabgaben entspricht, oder ob er darüber oder darunter liegt. Somit bekommt man Lösungshinweise auf die Frage, wer von der Geldablösung der Naturalabgaben profitiert hat: der Empfänger der Abgabe oder der Abgabepflichtige.

<sup>73</sup> Für die Punkte 10. bis 12. einige erklärende Stichworte. Im Falle eines Geldgrundzinses ist die Abgabenhöhe dem Urbareintrag zu entnehmen. Falls es sich um eine Naturalabgabe handelt, ist ihr Geldwert in den entsprechenden Tabellen abzulesen, die dem Urbar im Vorwort beigegeben sind. (Falls diese Tabellen fehlen, muß man auf die Wertangaben verzichten, wenn sie sich nicht aus anderen Quellen ergänzen lassen.) Die gleichzeitige Aufnahme von Naturalien und deren Wert erleichtert die Computerauswertung bzw. erhöht die Variationsbreite. Man kann sich auf diese Weise (mit Hilfe des Fensterprogramms) den Gesamtwert aller Abgaben berechnen oder z.B. auch die Gesamtanzahl aller Hühner ermitteln lassen. Sowohl computertechnisch als auch für das Lesen der codierten Quelle erweist sich auch hier wieder eine Sondernummer als zweckmäßig. (Sondernummer "001" bedeutet z.B., daß nur Wertangaben folgen, Sondernummer "002" besagt, daß nach der Wertangabe noch eine Stückzahl (wie z.B. 10.0 Stück Eier) folgt.) Die Stück-, Pfund- oder Malterangabe ist aus Gründen der verarbeitungstechnischen Vereinfachung immer als Kommazahl zu notieren, also "10.0" Stück Eier, statt "10" Eier.



### c) 3. Kategorie: Zwischen- und Endsummen

Im Urbar werden häufig Summen ausgeworfen. Diese wurden in früheren Urbarauswertungen ebenfalls verschlüsselt. Da diese Zwischen- und Endsummen jedoch vom Rechner ebenfalls ermittelt werden, ist ein Codieren dieser Kategorie nicht notwendig, für Kontrollzwecke sind diese Informationen jedoch nützlich.

### d) 4. Kategorie: Anlage

Unter "Anlage" ist (im Gegensatz zu einer Vermögenssteuer) eine ertragsabhängige Reichssteuer zu verstehen, die sich aus dem Steuerwert der jeweiligen Parzelle berechnen läßt<sup>74</sup>.

1. Die Besitzrechtsform (Spalten 32 bis 34).
2. Rechtliche Bodennutzung (Spalten 36 und 37).
3. Tatsächliche Bodennutzung (Spalten 39 bis 41).
4. Sondernummer für die Bearbeitung (Spalten 47 bis 49).
5. Größenangabe in Jauchert (Spalten 51 bis 54), Vierteln (Spalte 56) und Ruten (Spalten 58 bis 60)<sup>75</sup>.
6. Steuerwertangabe in Gulden (Spalten 62 bis 66), Kreuzern (Spalten 68 und 69) und Hellern (Spalte 71)<sup>76</sup>.
7. Anlage in Gulden (Spalten 73 bis 75), Kreuzern (Spalten 77 und 78) und Hellern (Spalte 80).

### e) Weitere Kategorien

Die quantifizierbaren Informationen des Gottmadinger Urbars von 1761 lassen sich auf die beschriebenen Kategorien aufteilen. Falls in anderen Urbaren weitere Informationen auftauchen, können zusätzliche Kategorien definiert und analog behandelt werden.

Es ist ebenfalls möglich, daß sich bei anderen neuzeitlichen Urbaren die Informationen an einzelnen Stellen von den hier behandelten Beispielen unterscheiden. Dann muß der Codeplan entsprechend modifiziert werden.

---

<sup>74</sup> Einige der ausgewerteten Urbare verdanken ihr Entstehen der Theresianischen Steuerreform (vgl. 2.1., Anm.14). Berechnungsgrundlage: Die Anlage beträgt jährlich für jede Parzelle zwei Heller pro Gulden ihres Wertes, also 1/240 des Wertes (vgl. Einleitung des Gottmadinger Urbars von 1761).

<sup>75</sup> Die Rutenangabe wurde gegebenenfalls gerundet.

<sup>76</sup> Die Hellerangaben bei Steuerwert und Anlage (Nr. 7) wurden gegebenenfalls gerundet.

### 4.3. Abspeicherung der Daten

Nach dem Erstellen des Codeplans hat es sich als zweckmäßig herausgestellt, das Urbar in einem ersten Schritt in codierter Form auf Listen zu übertragen und im zweiten Schritt verschlüsselt dem Rechner verfügbar zu machen. Beim codierenden Abschreiben ist zu empfehlen, für jede Kategorie gleichzeitig separate Listen zu führen. Das erleichtert die späteren Arbeiten entscheidend.

Die Informationen eines Urbars können über Lochkarten, Sichtbelege oder direkt am Bildschirm eingegeben werden. Bei den ersten vom Verfasser durchgeführten Urbarauswertungen wurden die Daten auf Lochkarten abgeleht. In einer Weiterentwicklung werden die Daten jetzt am Bildschirm eingegeben. Dieser Vorgang wird durch begleitende Editorprogramme<sup>77</sup> erleichtert und unterstützt<sup>78</sup>. Die gewählte Vorgehensweise wird anhand der Kategorie "Parzellen" beschrieben. Es gelten folgende Prinzipien:

1. Die Detailinformationen innerhalb eines Datensatzes werden zunächst in exakt derselben Reihenfolge verschlüsselt, wie sie in der Quelle stehen. Der Rechner ordnet in einem späteren Verarbeitungsschritt die Informationen innerhalb jedes Datensatzes so um, wie es für die Auswertung gewünscht wird.
2. Es werden immer nur diejenigen Detailinformationen innerhalb eines Datensatzes neu aufgenommen, die sich gegenüber dem vorhergehenden Datensatz ändern. Alle gleichbleibenden Informationen werden mittels Computer "nachbesetzt".

Die Eingabedatensätze müssen zunächst nicht spaltenweise aufgebaut sein. Sie enthalten jedoch in einer vorgegebenen Reihenfolge, die der Reihenfolge im Urbar entspricht, und in einem festen Format die verschlüsselten Informationen. Alle Informationen sind durch einen Punkt, der als Tabulator fungiert, getrennt. Falls eine Information im aktuellen und im vorhergehenden Urbareintrag identisch ist (z.B. die tatsächliche Nutzung der Parzelle), so braucht sie nicht erneut eingegeben zu werden, sondern es folgt sofort der nächste Trennpunkt. Über die Tabulatorenpunkte werden dann die Eingabedatensätze in Spaltenform gebracht.

---

<sup>77</sup> Für wertvolle und wiederholte Hilfestellungen beim Schreiben dieser Editorprogramme ist der Verfasser Herrn Dipl. Inf. *Wolfgang Mehl* zu Dank verpflichtet.

<sup>78</sup> Über Tabulatoren wird die formatierte Eingabe am Bildschirm gesteuert. Ausserdem werden immer nur diejenigen Informationen eingegeben, die sich gegenüber dem vorhergehenden Eingabedatensatz ändern. In einem weiteren Arbeitsgang trägt der Rechner die sich wiederholenden Informationen nach.

Es gibt aber auch Variablen, deren Inhalt sich von Urbareintrag zu Urbareintrag nur "relativ selten" ändert, wie z.B. der Herkunftsort der Besitzer. Solche Variablen werden im folgenden als "Sternchenvariablen" bezeichnet<sup>79</sup>. Bei den Eingabedatensätzen werden die Informationen der Sternchenvariablen weggelassen. Wenn sich der Inhalt einer Sternchenvariablen ändert (wenn also z.B. der Herkunftsort der Parzellenbesitzer wechselt), wird unmittelbar vor dem entsprechenden Datensatz der neue Wert der Sternchenvariablen angegeben. Diese Deklaration gilt dann für alle folgenden Datensätze bis zur nächsten Änderung. Vor dem ersten Datensatz muß allen Sternchenvariablen ein Wert zugewiesen werden.

Es werden beispielsweise die in Schema 6 wiedergegebenen Daten eingegeben<sup>80</sup>. Aus dieser Eingabe entsteht dann Schema 7<sup>81</sup>. Nun werden die "Sternchenvariablen" in die Datensätze eingefügt und alle Variablen in die endgültige Reihenfolge gebracht. Damit ergibt sich Schema 8. In einem dritten Schritt müssen nun die "leeren" Variablen nachbesetzt werden. Sie erhalten den Wert aus der jeweils darüber stehenden Zeile. Als Endergebnis ergibt sich Schema 9.

In diesem Zustand können die Eingabedaten durch das Fensterprogramm weiterverarbeitet werden. Es hat sich jedoch als zweckmäßig herausgestellt, zuerst die Größen-, Steuerklassifizierungs- und Steuerwertangaben in Realzahlen<sup>82</sup> unzuwandeln.

#### 4.4. Fehlerkontrolle

Dem in 3.1.3. beschriebenen Fensterprogramm kommt nicht nur eine zentrale Funktion bei der Urbarauswertung zu, sondern es wird auch zur Fehlerkontrolle eingesetzt. Mit seiner Hilfe können alle vorkommenden Ausprägungen der Variablen aufgelistet und danach einer Plausibilitätskontrolle unterzogen werden.

---

<sup>79</sup> Welche Variable im Einzelfall als Sternchenvariable behandelt wird, muß für jedes Urbar in Abhängigkeit von dessen Aufbau und Inhalt neu entschieden werden.

<sup>80</sup> Kommentare, die nicht zur Eingabe gehören, sind hier in Großbuchstaben geschrieben.

<sup>81</sup> Aus Gründen der Arbeitsökonomie wurden für die Größenangabe in Jauchert nur eine Spalte und für die Wertangabe in Gulden zunächst nur zwei Spalten vorgesehen. Tauchen im Urbar Angaben auf, die nicht in dieses Schema passen, so wird die entsprechende Information zunächst mit Hilfe der Sondernummer verschlüsselt.

<sup>82</sup> Das kann mit Hilfe kleiner Computerprogramme oder z.B. auch mittels SPSS geschehen.

\*zahl=11        (URBARKENNUNG)  
 \*datei=01       (KATEGORIE GRUNDSTÜCKE)  
 \*nna=0130       (NACHNAME DES BESITZERS)  
 \*vna=007        (VORNAME DES BESITZERS)  
 \*ber=000        (WEITERE SPEZIFIKATION ZUM BESITZER)  
 \*herk=310       (HERKUNFTSORT DES BESITZERS)  
 \*bes=00         (BESONDERHEITEN IN BEZUG AUF DIE  
                  GRUNDSTÜCKSNUMMER)  
 \*rech=100       (BESITZRECHTSFORM)  
 \*nutz=10        (RECHTLICHE BODENNUTZUNG)  
 \*sond=001       (SONDERNUMMER)

Spalten										1	1		1	1	1	1	
	1	2	3	4		5	6	7	8	9	0	1	2	3	4	5	6

0001.01.153.000.4462.0.0.037.00.12.6.00.07.51.6.00  
 \*herk=310  
 0002.01.100.050.3899...035..27.6..16.11.2  
 .02..051.3978.....22....24  
 .03.....044...6..20..0

Bedeutung der Spalten:

- 1 Seite im Urbar, 2 laufende Nummer je Seite
- 3 Nutzung, 4 Lage, 5 Grundstücksnummer
- 6 Jauchert, 7 Viertel, 8 und 9 Ruten (als Realzahl)  
 (6 bis 9: Parzellengrösse)
- 10 Kreuzer, 11 und 12 Heller (als Realzahl)  
 (10 bis 12: Steuerklassifizierung)
- 13 Gulden 14 Kreuzer 15 und 16 Heller (als Realzahl)  
 (13 bis 16: Steuerwert).

Schema 6

#### 4.5. Clio: Alternative zur vorgeschlagenen Datenerfassung

Clio ist als datenbankorientiertes Programmsystem zu bezeichnen, das die quellennahe Umsetzung eines Urbars in maschinenlesbare Form und das vergleichende Auswerten verschiedener Urbarkategorien bzw. zusätzlicher Quellen sehr vereinfachen kann. Allerdings ersetzt Clio nicht die Funktion des Fensterprogramms und es steht bis jetzt nur an sehr wenigen Rechenzentren zur Verfügung. Auf dem Göttinger Workshop vom 15. bis 18. Juli

```

*zahl=11
*datei=01
*nna=0130
*vna=007
*ber=000
*herk=310
*bes=00
*rech=100
*nutz=10
*sond=001
0001 01 153 000 4462 0 0 037 00 12 6 00 07 51 6 00
*herk=310
0002 01 100 050 3899 0 0 035      27 6      16 11 2
      02      051 3978      22      24
      03      044      6      20      0

```

### Schema 7

```

110001010130007000310011004462000001015300100037.00126.0007516.00
110002010130007000320011003899000502010000100035.00276.0016112.00
11  0201300070003200110039780005120  001      22      24
11  03013000700032001100      00  20  001 044.00

```

### Schema 8

1985 wurde jedoch die Möglichkeit anvisiert, das Fensterprogramm als Baustein in Clio zu integrieren. Damit wäre die hier vorgestellte Methode der Auswertung frühneuzeitlicher Urbare in ein größeres System eingebettet.

## 5. Schlußbetrachtung

Um die Urbarforschung, die wichtige Erkenntnisse im Rahmen der Sozial- und Wirtschaftsgeschichte erhoffen läßt, von mittelalterlichen auf neuzeitliche Urbare ausdehnen zu können, bedarf es wegen des großen Umfangs neuzeitlicher Urbare einer maschinellen Auswertungsmethode. Wie eine solche Urbarauswertung mittels Computer vor sich gehen kann, wurde im vorstehenden Text skizziert. Verschiedene, nach dieser Methode vom Verfasser durchgeführte Auswertungen neuzeitlicher Urbare beweisen, daß die vorgeschlagene Vorgehensweise zum Ziel führt.

Spalten:		1	1	1	1	1	111		1	1	2	2	2	2
1	2	3	4	5	6	7	8	9	0	1	2	3	4	567
											8	9	0	1
											2	2	2	3

11001010130007000310011004462000001015300100037.00126.0007516.00  
 11002010130007000320011003899000502010000100036.00276.0016112.00  
 11002020130007000320011003978000512010000100036.00226.0016242.00  
 11002030130007000320011003978000512010000100044.00226.0020240.00

Bedeutung der Spalten:

- 1 URBARSCHLÜSSEL, 2 SEITE, 3 LAUFENDE NUMMER PRO SEITE;
- 4 NACHNAME, 5 VORNAME, 6 ZUSATZ, 7 HERKUNFTSORT;
- 8 DATEISCHLÜSSEL;
- 9 RECHTSFORM;
- 10 GRUNDSTÜCKSNUMMER, 11 BESONDERHEITEN B. D. GRUNDSTÜCKSNUMMER;
- 12 LAGE, 13 RECHTL. NUTZUNG, 14 TATS. NUTZUNG, 15 SONDERNUMMER;
- 16 JAUCHERT, 17 VIERTEL, 18 RUTEN;  
(16 bis 18: Größe der Parzelle)
- 19 KREUZER, 20 HELLER;  
(19 und 20: Steuerklassifizierung)
- 21 GULDEN, 22 KREUZER, 23 HELLER  
(21 bis 23: Steuerwert)

### Schema 9

Der Kernpunkt der Methode ist das Aufteilen der Urbareinträge in verschiedene Kategorien, das separate Verschlüsseln jedes Datensatzes und die Aufbereitung der Daten durch das Fensterprogramm<sup>83, 84</sup>

Damit können nun auch umfangreiche neuzeitliche Urbare, die bislang nur ansatzweise in die historische Forschung Eingang gefunden haben, umfassend quantifizierend ausgewertet werden.

<sup>83</sup> Das Fensterprogramm ist in Standard - Pascal geschrieben und "läuft" auf jedem Computer mit Pascal - Compiler. Der Jobrahmen muß allerdings an die jeweilige Rechneranlage angepaßt werden. Es stammt in der hier abgedruckten Version von Dipl. Phys. Otto Stolz, Konstanz. Vgl. dazu Anm. 36.

<sup>84</sup> Aus Platzmangel mußte auf die Wiedergabe des Programmes im Druck verzichtet werden.

## Literaturverzeichnis

- Beutel, M./Schubö, W.: Statistikprogrammsystem für die Sozialwissenschaften. 4. Aufl., Stuttgart/New York, 1983.
- Dollinger, P.: Der bayerische Bauernstand vom 9. bis zum 13. Jahrhundert, München, 1982.
- Dopsch, A.: Die Herausgabe von Quellen zur Agrargeschichte des Mittelalters, in: DtGgbll 6 (1905), S.145 - 167.
- Erbs, H.-E./Stolz, O.: Einführung in die Programmierung mit Pascal, 2. Aufl., 1984.
- Fossier, L.: Informatique et documents médiévaux, in: Annales. Economies, Sociétés, Civilisations 31 (1976), S.1131 - 1142.
- Göttmann, F./Rabe, H./Sieglerschmidt, J.: Regionale Transformation von Wirtschaft und Gesellschaft, in: Schr.VG Bodensee 102 (1984), S. 115 - 173.
- Inama - Sternegg, K. T. von: Abhandlungen über Quellen der deutschen Wirtschaftsgeschichte (=Sitzungsber. d. kais. Ak. d. Wiss., phil.-hist. Cl. 84), Wien, 1876.
- Inama - Sternegg, K. T. von: Über Urbaren und Urbarialaufzeichnungen, in: Archival. Zs. 2 (1877), S. 26 - 52.
- Kleiber, W.: Urbare als sprachgeschichtliche Quelle. Möglichkeiten und Methoden der Auswertung (=Veröff. d. Komm. f. geschichtl. Landesk. in Baden - Württemberg, Reihe B, 33), Stuttgart, 1965.
- Kleiber, W.: Das Aufkommen der deutschen Sprache in domanialen Rechtsquellen (Urbaren) Südwestdeutschlands zwischen 1250 und 1450, in: Aleman. Jb. 1973/75, S. 202 - 220.
- Knuth, D.E.: The Art of Computer Programming, Bd.3: Sorting and Searching.
- Kropf, R.: Zur Auswertung von Urbaren mittels Computer, in: Mitt. d. Inst. f. Österr. Gesch. 77 (1969), S. 141 - 144.
- Lamprecht, K.: Deutsches Wirtschaftsleben im Mittelalter, 3 Bde., Aalen, 1960 (Erstveröff. 1885/86).
- Ott, H.: Studien zur spätmittelalterlichen Agrarverfassung im Oberrheingebiet (=Qu. u. Forsch. z. Agrargesch. 23), Stuttgart, 1970.
- Portmann, U.: Elektronische Datenverarbeitung und mittelalterliche Urbare. Auswertungen mit Urbaren des Stiftes Beromünster aus dem 14. Jahrhundert (Lizentiatenarbeit), Fribourg, 1977.
- Quarthal, F.: Landstände in Schwäbisch - Österreich (=Schr. z. Südwestdt. Landesk. 16), Stuttgart, 1980.
- Richter, G.: Kompetenzstreitigkeiten württembergischer Behörden im 18. Jahrhundert, in: Zs. f. Württ. Landesgesch. 27 (1968), S. 339 - 360.

*Richter, G.*: Bedeutung und Publikationsmöglichkeiten von Urbaren des 16. bis 18. Jahrhunderts. Überlegungen nach Stuttgarter Quellen, in: Zs. f. Württ. Landesgesch. 31 (1972), S. 232 - 248.

*Schrenk, C.*: Methoden der Auswertung frühneuzeitlicher Urbare anhand des Orsinger Urbars von 1758, In: *Göttmann, F./Rabe, H./Siegler Schmidt, J.*: Regionale Transformation von Wirtschaft und Gesellschaft, in: Schr.VG Bodensee 102 (1984), S. 115 - 173. Hier: S. 153 - 163.

*Stenzel, I.*: Fehler und Fallen der Statistik, Bern, 1982.





**Ein Programm- und Methodenpaket  
zur Rekonstruktion von Klimaverhältnissen  
seit dem Hochmittelalter**

## **1. Einleitung**

Während Veränderungen der Klimaverhältnisse in die meisten globalen Prognosemodelle als Risikofaktoren einbezogen werden, hat die Geschichtsschreibung ihren Stellenwert in der vergangenen wirtschaftlichen und sozialen Entwicklung bisher nur zurückhaltend beurteilt. Verschiedene Historiker, so etwa *Braudel* (1949, 1979) und *Mols* (1979), stufen die Bedeutung von Klimaschwankungen für die demographische und konjunkturelle Geschichte der frühen Neuzeit zwar recht hoch ein. Sie sind aber nicht in der Lage, solche Einflüsse kontinuierlich aufzuzeigen und zu gewichten. Meteorologische Meßreihen, welche dies stringent erlauben würden, liegen in genügender Dichte nur für die letzten 125 Jahre vor.

Anhand schweizerischen Materials ist eine neue Methode entwickelt worden, um die Fülle von Wetternotizen und Umweltbeobachtungen und frühen Messungen, die in Schriftquellen wie Chroniken, Tagebüchern, Briefen, privaten und obrigkeitlichen Buchhaltungen enthalten sind, zu Schätzwerten von Temperatur und Niederschlag für die einzelnen Jahreszeiten und Kalendermonate zu verdichten (*Pfister*, 1984). Da sich allein für die Schweiz im Zeitraum 1525-1863 über 33.000 solche Beobachtungselemente gefunden haben, ist anzunehmen, daß das Quellenmaterial für analoge Studien in anderen Teile Europas ebenfalls ausreichen dürfte. Rekonstruktionen sind sogar für die Zeit des Hoch- und Spätmittelalters möglich, wie eine in diesem Jahr zu erwartende Arbeit von *Pierre Alexandre* (Liège) zeigt, allerdings mit erheblich geringerer zeitlicher und elementspezifischer Trennschärfe.

Im folgenden sollen das Verarbeitungskonzept und die Software vorgestellt werden. Es ist bis heute in einer Weise erweitert worden, daß es die Grundlage für eine europäische klima- und umweltgeschichtliche Datenbank abgeben könnte.

## **2. Zur Typologie klimageschichtlicher Daten aus Schriftquellen**

Die anfallenden Informationselemente lassen sich typologisch wie in Figur 1 gliedern. Wesentlich ist die Unterscheidung zwischen klimaspezifischen Daten und Proxy-Daten. Erstere beschreiben unmittelbar Elemente oder Manifestationen des Witterungsverlaufs, letztere beobachten Auswirkungen von klimatisch gesteuerten physikalischen und biologischen Prozessen, wie

etwa den Stand von Gewässern, die Eisbildung auf Seen, die Blütezeit von Kulturpflanzen oder Menge und Zuckergehalt des geernteten Weinmosts. Proxy-Daten lassen sich im Überlappungsbereich mit Instrumentenmessungen kalibrieren. Sie erlauben damit eine Grobschätzung von Temperatur- und Niederschlagsmustern. Der Vergleich mit den parallel dazu vorliegenden Witterungsberichten ermöglicht eine feinere zeitliche und elementspezifische Aufschlüsselung.

Die Kalibrierung von Proxy-Daten sollte vorgängig durchgeführt werden, ebenso wenn nötig die Homogenisierung von frühen Meßreihen. Dies fordert eine Zusammenarbeit mit Geografen, Meteorologen und Agronomen.

Im Verarbeitungskonzept wird zwischen zwei Gruppen von Datentypen unterschieden:

- ausschließlich deskriptive Informationen: sie werden ohne Vorinterpretation über einen numerischen Code erfaßt.
- "numerischen Informationselementen": dazu gehören quantifizierte deskriptive Informationen wie etwa Angaben über die Häufigkeit von Niederschlagstagen, wie sie aus Witterungstagebüchern gewonnen werden können, im weiteren biologische und physikalische Proxy-Daten, sowie alte Meßreihen. Sie werden mit gängigen Software-Paketen SAS, SPSSX) oder dem Programm METIF (zur Anwendung von täglichen Beobachtungen) vorinterpretiert und anschließend in standardisierter Form angegeben.

### **3. Aufbau von Programm und Codebuch<sup>1</sup>**

#### **3.1. Aufbau der Records**

Die Kolonnen 1-14 aller Records bleiben den chronologischen, lokalen und quellspezifischen Attributen (Name des Beobachters, Quellenverweis) vorbehalten. Grundelement des Zeitrasters ist die Dekade; ein Record entspricht immer mindestens einer Zehntagesperiode; er kann aber auch größere Zeiteinheiten ausdrücken. Die Wahl der Zehntagesperiode stellt eine zeitliche Auflösung von mehr als einem Monat sicher, wie sie für die Interpretation von demographischen und wirtschaftsgeschichtlichen Daten erforderlich ist, ohne daß die Datei mit einer Flut von täglichen Beobachtungen überladen wird. Die Kolonne 6 drückt zusätzlich zur Dekade den Datentyp aus und erhält damit auch eine programmsteuernde Funktion. Ist ihr Wert gleich oder kleiner 4, so enthält der betreffende Record deskriptive, nicht quantifizierbare Daten mit einer Aufschlüsselung nach Dekaden oder größeren Zeiteinheiten.

---

<sup>1</sup> Programm-Ablauf: Figur 2, Codebuch-Übersicht: Figur 3

Ist der Wert größer, enthält der Record Proxy-Daten oder Häufigkeitswerte aus Witterungstagebüchern (Output des Programms METIF).

Falls tägliche Beobachtungen vorliegen, werden sie in die Kolonnen 16-54 aufgenommen. (In diesem Falle durchläuft der Record vorgängig das Zählprogramm METIF). Kolonne 56 steht für quellenkritische Attribute zur Verfügung. Witterungsbeobachtungen im Zeitraster von Dekaden, Monaten oder Jahreszeiten werden im Feldbereich 57-74 eingetragen. Auf Kolonne 75 wird angemerkt, ob eine Fußnote vorgesehen ist: Fußnoten sind für jene Beobachtungen konzipiert, die sich durch den Code nicht adäquat ausdrücken lassen (u.a. Beschreibungen von Katastrophenereignissen).

Bei der Proxy-Information wird der Datentyp durch die Kolonne 6 festgelegt:

- 5: Niederschlagsmessungen
- 6: Häufigkeitswerte aus Witterungstagebüchern
- 7: Temperaturmessungen
- 8: für Monat Juni: Getreide-Zehntsteigerungsdaten  
für Monat Juli: Quantitative Angaben über Weinmosterträge  
für Monat September: Dichte des Spätholzsegments v. Baumringen  
(Ein Indikator für Temperaturen Juli-September, der seit dem frühen Mittelalter zur Verfügung steht).
- 9: Für das Resultat der Interpretation, die Temperatur- u. Feuchtigkeits-Indices

Weitere neu auftauchende Proxy-Datentypen könnten durch alphanumerische Zeichen ausgedrückt werden.

### **3.2. Die vorverarbeitenden Programme: METIF, SEPCLIM, SORT**

Die Aufbereitung und Kalibrierung von Proxy-Daten erfolgt zweckdienlich durch vorhandene *Statistik-Software* (SAS, SPSSX).

Mit dem *PL-1 Programm METIF* werden tägliche Beobachtungen aus Witterungstagebüchern ausgezählt. Sie ordnen sich in den Feldbereich 16-54 ein. Die Kolonnen 16-18, 20-22, ..., 52-54 bezeichnen den ersten, zweiten, respektive zehnten Tag der entsprechenden Dekade. Jedem Tag sind je eine Kolonne für Winde, Thermik und Witterungscharakter zugeordnet. Der Papierausdruck liefert Häufigkeitstabellen der verschiedenen Ereignisse, was eine Kontrolle erlaubt. Auf ein weiteres File werden die monatlichen Häufigkeitswerte in einer Form ausgegeben, die ein unmittelbares Einlesen in

das dem Hauptprogramm SORTCLIM vorgelagerte IMB-Sortierprogramm erlaubt (Kennzeichnung der Dekade mit 6). Nach den Kolonnen 1-14 mit den üblichen Angaben folgen von Kolonne 16 an 29 zweistellige Ziffern, welche die Häufigkeit von Tagen mit Niederschlag, Schneefall, Schneebedeckung, durchgehender Bewölkung etc. ausdrücken.

Im *PL-1 Programm SEPCLIM* werden alle mit  $\text{Dec} \leq 4$  codierten Daten vorverarbeitet. SEPCLIM hat verschiedene Funktionen:

1. Der Bearbeiter codiert wenn möglich alle für eine Dekade benötigten Attribute wie Chronologie, Beobachtungsort und Quelle auf einem einzigen Record. Damit erspart er sich eine mehrfache Eingabe dieser Attribute. SEPCLIM trennt nun diese kompakten Eingabe-Records in der Weise auf, daß jede Position – z.B. Naturerscheinungen, pflanzenphänologische Beobachtungen etc. – auf einem eigenen Record ausgegeben wird.
2. Das Programm kann Quellen aus beliebigen Ländern verarbeiten. Zur eindeutigen Kennzeichnung der Quellennummer setzt der jeweilige Bearbeiter eine durch Konvention festgelegte Landesnummer ein. Diese erscheint im Output in Kolonne 16-18, wodurch alle anderen Beobachtungen um 4 Stellen nach rechts verschoben werden.
3. Das Programm kontrolliert die Daten auf gewisse Eingabefehler:
  - Beobachtungen müssen zwischen den Jahren 1000 und 1980 liegen
  - nichtbesetzte Positionen des Codebuches werden als solche angezeigt
  - elementare logische Kontrollen: so können die Kirschen beispielsweise in der nördlichen Hemisphäre kaum im Dezember blühen.
4. Die Kennzahlen für Monate und Jahreszeiten, die vom Beobachter nach dem gängigen Schema der Monatszahlen (1-12, anschließend Jahreszeiten) codiert werden, müssen temporär umgruppiert werden, damit bei der Ausgabe eine logische Ordnung entsteht, welche die jahreszeitlichen Beobachtungen auf die zugehörige Gruppe von Kalendermonaten folgen läßt (z.B. Angaben für den gesamten Winter nach jenen für den Februar).

Ein *IBM-Sortierprogramm* nimmt alle auf verschiedenen Eingabefiles vorhandenen Datentypen auf und sortiert sie chronologisch, räumlich (Land, Region) und nach Datentypen.

### 3.3. Programme SORTCLIM-Test und SORTCLIM-Def

Alle im SORT-Pool zusammengefaßten Daten werden zunächst ins Testprogramm SORTCLIM eingelesen. Die Testvariante entspricht praktisch der definitiven Variante von SORTCLIM. Sie dient lediglich der Eruierung von allfälligen Fehldaten und bereitet das Anmerkungsfile (ANM) vor, und zwar mit den Angaben von Jahr (Kol. 1-4), Monate (6-8), Dekade (10) und Quelle

(15-18). Um Fehler zu finden, muß in die mit SEPCIM verarbeiteten Daten, möglicherweise in die Ursprungsdaten zurückgegriffen werden. Erst wenn die Daten den Testlauf bestanden haben, durchlaufen sie das definitive SORTCLIM.

### **3.3.1. Deklarationen, Einlesen von Schwellenwerten**

Neben den sechs Ein- und den zwei Ausgabefiles werden Bereichsvariable und gewöhnliche Variable deklariert. Es folgen ON-Bedingungen, wobei das File ANM als Input für die später folgenden Anmerkungen kurzfristig geöffnet und dann wieder geschlossen wird.

Im nächsten Schritt werden vier externe, d.h. nicht durch die Vorprogramme gelaufene Files geöffnet, mit dem Befehl GET eingelesen und wieder geschlossen. Diese Files erfüllen die Funktionen von *Vergleichswerten*. Es sind dies die Files REF, TMP, NIL und PO. Die Daten werden hier durch Kommas voneinander getrennt erfaßt. Das File REF enthält: Quelle, Region, Ortsnamen, Stationshöhe, Anzahl Buchstaben für den Ortsnamen, Autorennamen, Quellennummer, heutige Vergleichsstationen für die Interpretation von Niederschlagsmessungen und -häufigkeit, Land, Kalendertyp, Lebensdaten des Verfassers der Quelle. Die übrigen externen Files beinhalten Schwellenwerte von Temperaturen (TMP), Niederschlagstagen (NIL) und Niederschlagsmessungen (PO) an Vergleichsstationen im heutigen meteorologischen Meßnetz aus der Meßperiode 1901-1960. Das Einleseprozedere umfaßt je fünf Schwellenwerte für die 12 Monate des Jahres für eine beliebige erweiterbare Anzahl von Vergleichsstationen. Die Berechnung der fünf Datenwerte erfolgt durch Ordnung in z.B. aufsteigender Reihenfolge aller 60 Januar-Durchschnittstemperaturen des Zeitraumes 1901-1960. Die Verteilung wird nachher in Gruppen von gleicher Größe eingeteilt. Wird die Reihe z.B. auf 12 Gruppen verteilt, bezeichnet man deren Grenzwerte, häufig aber auch die Gruppe selbst, als Duodezile. 50 % aller Werte liegen unterhalb des neunten und oberhalb des dritten Duodezils ("die mittlere Hälfte"), je 17 % zwischen dem neunten und dem elften, respektive zwischen dem dritten und dem ersten Duodezil; je 8,5 % liegen oberhalb des elften, respektive unterhalb des ersten Duodezils.

### **3.3.2. Einlesen von Steuerdaten**

Jetzt erfolgt die Öffnung des Files DAT, welches die Ziffern wieder in Sprache umsetzt, jetzt aber chronologisch geordnet. Darin sind alle mit den Vorprogrammen verarbeiteten Daten zu finden, d.h. die Datentypen mit den Ziffern 0 bis 9 auf Kolonne 6. Eingelesen werden in einem ersten Schritt

die Ziffern für Jahr, Monat, Dekade, Region, Quelle und, fakultativ, Landesnummer. In einem zweiten erfolgen die Beobachtungszuweisungen, d.h. die Umsetzung der Ziffern in lesbare Sprache.

Zugewiesen werden:

- a) die Monatsnamen, z.B. "IF Monat = 9 then MNAME (Monatsname) = SEPTEMBER", und
- b) die Beobachtungsorte; hier erfolgt nun ein Vergleich zwischen dem externen File REF und dem File DAT, gewissermaßen als Kontrolle. Entsprechen sich Region und Quelle sowie eventuell das Land der beiden Files, wird der Beobachtungsort des Files REF ausgedruckt. Entsprechen sich Quelle und Land, nicht aber die Region, wird die gesuchte Region zugewiesen; z.B. REG = 13 / Regionsname = Mittelbünden.

An dieser Stelle sei nochmals darauf hingewiesen, daß mit wenig Arbeit das Programm wirkungsvoll ausgebaut werden kann, je nach der Quellenlage.

Zum richtigen Ausdrucken der Steuerdaten bedarf es anschließend genauer Anweisungen für Jahr, Monate und Dekaden (x, y, z.) Die Komponenten x, y, z sind variabel; für alle Fälle sind Ausdrücke vorgesehen:

Beispiele:

- Angaben für Monatsdaten ohne Dekadenangabe (Januar, Sommer, etc.)
- Dekadenwechsel mit Druck der Dekade
- Dekadenwechsel mit Druck "gesamter Monat"

Die beschriebenen Zuweisungs- und Ausdruckbefehle sind einmalig und gelten für alle folgenden DAT-Schritte als verbindlich.

### **3.3.3. Hauptverzweigungen, Einlesen und Zuweisung**

Die Einweisung des Programms in die einzelnen Datentypen ist das zentrale Steuerungssystem. Von hier aus werden mit Sprungbefehlen die Datentypen ausgewählt; die Sprunganweisungen lauten:

```
IF DEC (Kol. 6) = 4 THEN GOTO RV
IF DEC (Kol. 6) = 5 THEN GOTO MD
IF DEC (Kol. 6) = 6 THEN GOTO FC
IF DEC (Kol. 6) = 7 THEN GOTO TP
IF DEC (Kol. 6) = 8 THEN GOTO PX
IF DEC (Kol. 6) = 9 THEN GOTO IND
```

Die Sprunganweisungen werden nun im folgenden in der Programmreihenfolge erläutert:

## 1. RV ( $\text{DEC} \leq 4$ ); qualitative und quantifizierte Beobachtungen

Auf den Kolonnen 56-75 (60-79) erfolgt das Einlesen der Daten mit dem Befehl GET FILE (DAT) EDIT. Benötigt werden die Parameter Interpretationsstufe der Quelle, Winde, Thermik, Witterungscharakter, Schneedecke, besondere Naturereignisse, Tagescode für Phänologie, phänologische Beobachtungen, witterungsbedingte Schäden, Ernteergebnisse und der Anmerkungsverweis.

Im zweiten Schritt müssen den codierten Ziffern die wortmäßigen Beobachtungsinhalte zugeordnet werden. Die Zuweisungen erfolgen nach ganzen Monaten und Dekaden.

Beispiel:

WETH 1 = 6 (Kolonne 62, resp. 66) und WETH 2 = 9 (Kolonne 63, resp. 67) ergibt die Gesamtziffer 69, die Zuweisung lautet: "Vereinzelte Aufhellungen". Selbstverständlich wird gleichzeitig für die Beobachtung der entsprechende Platz reserviert, hier also 17 Positionen. Zum Schluß erfolgen die Ausdrucksanweisungen der Beobachtungen, wie sie auf dem Ausgabefile ASA erscheinen.

## 2. MD ( $\text{DEC} = 5$ ); Niederschlagsmessungen

Nach dem Einlesen der "historischen" Niederschlagsmessungen (Kol. 16-18) wird nun der Vergleich mit der Normalperiode 1901-1960 hergestellt (File PO). Wenn möglich erfolgt der Vergleich mit der Station des gleichen Beobachtungsortes, sonst mit der nächstgelegenen (vergleichbaren) Station des meteorologischen Meßnetzes. Die Zuordnungen der alten Niederschlagsmessungen zur alten Normalperiode lauten:

- normal: Alle Monatswerte liegen unterhalb des neunten Duodezils und oberhalb des dritten Duodezils.
- nass/trocken: Alle Monatswerte liegen zwischen dem neunten und dem elften Duodezil, respektive zwischen dem dritten und dem ersten Duodezil.
- extrem nass/trocken: Alle Monatswerte, die oberhalb des elften oder unterhalb des ersten Duodezils liegen.

Beim Ausdruck wird zusätzlich das Mittel aufgeführt, d.h. der Durchschnitt der Periode 1901-1960.



### 3. FC (DEC = 6); Häufigkeitsangaben aus Witterungstagebüchern

Eingelesen werden die Daten so, wie sie vom Programm METIF aufgearbeitet worden sind: von Kolonne 16 an 29 zweistellige Ziffern mit der Kolonne 75 als Abschluß.

Anschließend erfolgt die Zuweisung zu den Niederschlagstagen aus Witterungstagebüchern im Vergleich mit dem File NIL (Niederschlagstage in mm). Wiederum erfolgt die Qualifizierung aufgrund des Ranges innerhalb der Duodezilverteilung: extrem naß, naß, veränderlich (= normal), trocken, extrem trocken.

Beim Druck wird zusätzlich M ausgedruckt, der Mittelwert der Periode 1901-1960. Die Zuweisungen der *Himmelsbedeckung* am gesamten Monat lauten:

- bewölkt
- sonnig
- Morgennebel – sonnig
- Nebel

“Sonnig” erscheint, wenn die Zahl der sonnigen Tage mehr als 50 % beträgt. “Morgennebel – sonnig”, bzw. “Nebel” erscheint, wenn die Zahl der Tage mit Morgennebel und anschließender Auflösung, bzw. der durchaus nebligen Tage mehr als 50 % beträgt.

Die Zuweisung der *Temperaturangaben* sind “warm” und “kalt”. Erstes erscheint, wenn die Zahl der vom Beobachter als “warm” bezeichneten Tage größer als 12 und um mindestens 35 % größer als die der von ihm als “kalt” bezeichneten Tage ist. “Kalt” erscheint, wenn die Zahl der vom Beobachter als “kalt” bezeichneten Tage größer als 12 und um mindestens 35 % größer als die der von ihm als “warm” bezeichneten Tage ist.

Die *Schneebedeckung* wird mit verschiedenen Attributen ausgedrückt:

- sehr lange Schneebedeckung (Schneedeckentage > 25; Monate Dez., Jan., Feb.)
- lange Schneebedeckung (> 12; Monate Okt., Nov., März, April, Mai)
- sehr geringe Schneebedeckung (<6; Monate Dez., Jan., Feb.)
- kein Schnee

Weiters erfolgen Zuweisungen zu Gewitterhäufigkeit (“gewitterhaft”) und Windrichtungen (Bise, NO-O-Wind; Föhn, SO-O-Wind; W-SO-Wind; NW-N-Wind).

#### 4. TP (DEC = 7); Temperaturmessungen

Das Verfahren verläuft äquivalent dem 2. Punkt (Niederschlagsmessungen). Die Zuordnungen in Grad Celsius lauten nun:

- extrem warm
- warm
- normal
- kalt
- extrem kalt

#### 5. PX (DEC = 8); Proxy-Daten

Einlesen, Zuweisung und Ausdruck erfolgen für:

- a) *Getreide-Zehntsteigerungsdaten*: Eingelassen werden mit Attribut MON (Monat) = 6 (Juni) auf Kolonne 16-18 die Schwellenwerte der durchschnittlichen Abweichungen sämtlicher verfügbaren Serien: Getreide-zehntsteigerungen

- sehr früh:	Abweichung vom Mittel	> 8 Tage
- früh:	"	8-7 Tage
- etwas verfrüht:	"	6-5 Tage
- normal:	"	+4 bis -4 Tage
- etwas verspätet:	"	-6 bis -5 Tage
- spät:	"	-8 bis -7 Tage
- sehr spät:	"	< 8 Tage

Vor der klimageschichtlichen Interpretation sind die vorhandenen Serien zu homogenisieren (vgl. Pfister, 1984: 91 f.).

- b) *Weinetragsdaten*: Einlesen der Mittel aus mehreren Serien (Kolonnen 16-19, Monat = 7), wobei die Abweichung in % des Trends angegeben wird (vgl. Pfister, 1984, 95 f.). Die Schwellenwerte werden definiert:

> 100 %	Weinmosterträge	sehr groß
25 - 99 %:	"	groß
- 25 bis + 24 %:	"	normal
- 26 bis - 51 %:	"	gering
< 51 %:	"	sehr gering

Aussagekräftig sind diese Daten für den Monat Juli.

- c) *Spätholzdichte*: Eingelassen werden dendroklimatische Daten (Kolonnen 16-17; Monat = 9), für die Schweiz z.B. gewonnen aus dem bernischen Lauenen. Die Zuordnung der Schwellenwerte lautet:

< 880	Spätholzdichte	sehr gering
880 - 919	"	gering
920 - 959	"	unternormal
960 - 1039	"	normal
1040 - 1079	"	übernormal
1080 - 1119	"	groß
> = 1120	"	sehr groß

Besondere Bedeutung erlangen diese Daten für den Monat September (vgl. Pfister, 1984: 97).

## 6. IND (DEC = 9); Temperatur- und Niederschlagsindices

Eingelesen auf den Kolonnen 16-19. Um einen eigentlichen Datentyp handelt es sich hier nicht. Die Inhalte der Wortquellen – wenn möglich – jedes einzelnen Monats werden durch eine numerische Skala ersetzt. Die Qualität und Dichte der schweizerischen Daten hat es erlaubt, Indices für einzelne Monate zu deduzieren und nach Temperatur- und Niederschlagsverhältnissen zu differenzieren. Die Kalibrierung der Indices richtet sich an der Duodezilverteilung für den Zeitraum 1901-1960 aus. Die Indices (7 Stufen) reichen von:

- 3 (extrem kalt, extrem trocken) bis

+ 3 (extrem warm, extrem naß);

0 steht für normale Monate nach Maßgabe der Meßperiode 1901-1960.

Für die vorinstrumentelle Periode wurden die Verhältnisse in dieser Weise geschätzt, sofern die Qualität des Materials diesen Grad der Differenzierung erlaubte. Wo dies nicht der Fall war, wurde nur zwischen drei Stufen unterschieden (+1 / 0 / -1).

### 3.3.4. Print

Die Druckanordnungen sehen zuerst die Zuweisung der *Anmerkungen* vor. Die Anmerkungs-texte sind auf einem speziellen File gespeichert, es geht hier darum, hinter eine Beobachtung, die einer näheren Erläuterung bedarf, eine Anmerkungsnummer zu setzen. Theoretisch kann jede Beobachtung eine Anmerkung enthalten. Tatsächlich wird aber nur ein Teil aller Beobachtungen eine zusätzliche Beschreibung brauchen. Die Testvariante vom Programm SORTCLIM druckt ein File ANM aus, das die Attribute aller beim Codieren für eine Anmerkung vorgesehenen Records enthält. Die tatsächliche Zahl der ausformulierten Anmerkungen wird in der Regel kleiner ausfallen. Wird eine vorgesehene Anmerkung nicht realisiert, so wird der entsprechende Record

auf dem File ANM gelöscht. Andererseits können zusätzliche Anmerkungen eingefügt werden, sofern auf dem entsprechenden Record im File DAT auf Kolonne 75 der Anmerkungsimpuls nachgetragen wird. Die definitive Variante von SORTCLIM vergleicht die potentiellen Anmerkungen (Kolonne 75) mit den tatsächlich realisierten (File ANM). Eine Anmerkungsnummer wird nur beigefügt, wenn die Attribute der beiden Records übereinstimmen. Außerdem wird in diesem Falle ein weiteres (Kontroll-)File mit den Attributen der definitiv realisierten Anmerkungen ausgegeben.

Die Daten werden schon bei der Codierung nach dem Gregorianischen Kalender erfaßt. Um diese Umrechnungen kenntlich zu machen, wird der verwendete Kalendertyp – der bei den Attributen der betreffenden Quelle im File REF vermerkt ist – dem Ausdruck des entsprechenden Records beigefügt.

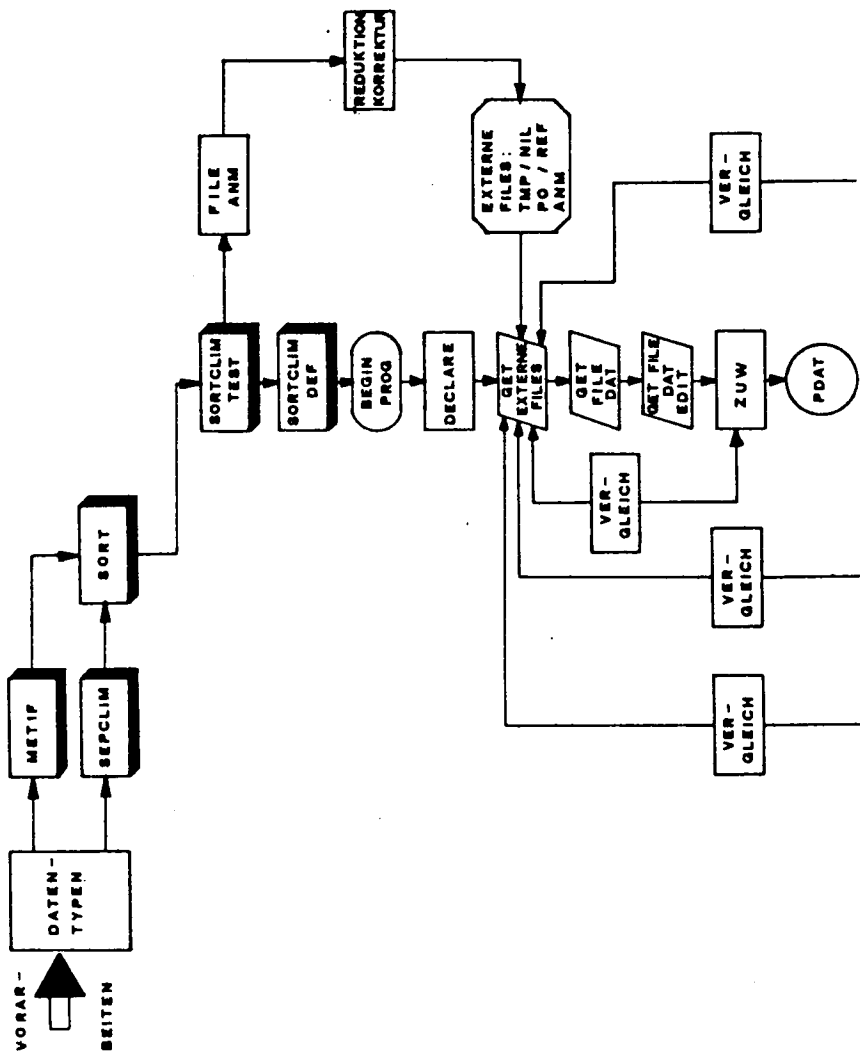
Die beiden letzten Schritte des Programms ordnen die Zeilenlängen für die Beobachtungsausdrucke auf dem Ausgabefile ASA und steuern die Fehlermeldungen.

#### **4. Mehrsprachigkeit und Ausbaufähigkeit**

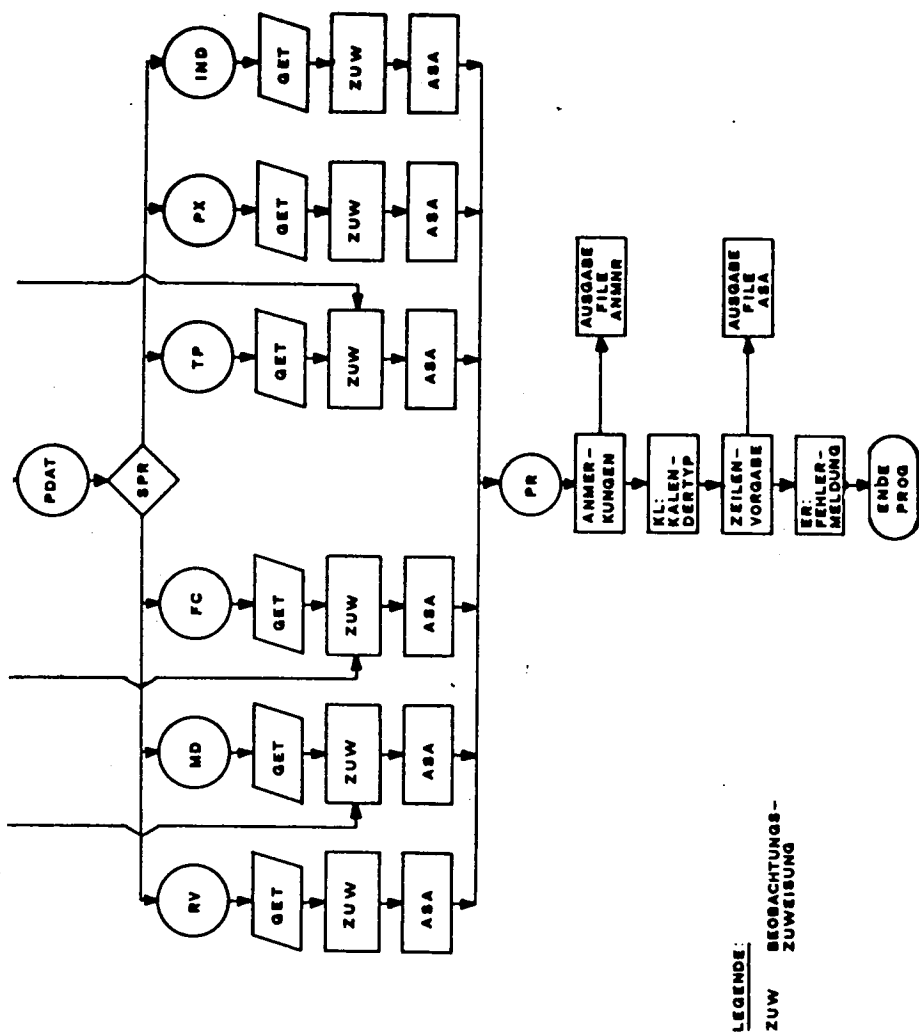
Die Erfassung der Information über einen numerischen Code und die anschließende Zuordnung von Wortelementen erlauben es, mit verhältnismäßig geringem Aufwand Programmvarianten für weitere Sprachen zu entwickeln. Die Witterungsdatei CLIMHIST (Fig. 4), welche die Schweizer Daten enthält, ist bereits auch in englischer, französischer (Fig. 5) und italienischer Sprache erhältlich. Damit ist eine – unseres Erachtens wesentliche – Voraussetzung für die supranationale Brauchbarkeit der vorgestellten Software gegeben. Weitere Voraussetzungen – größtmögliche Flexibilität und Ausbaufähigkeit – sind von der Programmstruktur her gegeben. Ob unser Ansatz in Richtung auf eine internationale klimageschichtliche Datenbank weiter verfolgt wird, dürfte vor allem von der Initiative interessierter Forscher in anderen Ländern abhängen. Falls ein Bedürfnis nach einem solchem Material vorhanden ist, erscheint uns eine Koordination allfälliger Bemühungen mit den bereits bestehenden Lösungsmöglichkeiten erwägenswert.

Genese Inhalt	Natürliche Daten	Anthropogene Daten		
		Schriftquellen		Sach- quellen
<b>Klimaspezifische Daten:</b>  Witterungsverlauf und Klimaelemente		Witterungsbeschreibungen	Instrumentelle Messdaten	
		<ul style="list-style-type: none"> <li>– Extremereignisse</li> <li>– Grundzüge des Witterungsverlaufs</li> <li>– tägliche Aufzeichnungen</li> </ul>	<ul style="list-style-type: none"> <li>– Luftdruck</li> <li>– Temperatur</li> <li>– Niederschlag</li> <li>– Pegel</li> </ul>	
<b>Indirekte Daten:</b> klimatisch beeinflusste Prozesse und Erscheinungen	<b>Physikal. Daten</b> <ul style="list-style-type: none"> <li>– Isotopen (<math>^{16}\text{O}/^{18}\text{O}</math>)</li> <li>– Moränen</li> <li>– Sedimente usw.</li> </ul>	<b>Physikalische Daten</b> <ul style="list-style-type: none"> <li>– Wasserstände</li> <li>– Schneefall</li> <li>– Schneebedeckung</li> <li>– Vereisung von Gewässern</li> </ul>	<b>Biologische Daten</b> <ul style="list-style-type: none"> <li>– Blüte- und Reifezeit von Pflanzen</li> <li>– Ertrag und Zuckergehalt von Weinmost</li> <li>– Zeitpunkt von Getreide-ernte und Weinlese</li> </ul>	alte Bild Darstellungen und Fotos; Pläne und Karten, Bauwerke, Siedlungen, Verkehrswege, aufgelassene Fluren und Kulturen, archaische Funde usw.
	<b>Biolog. Daten</b> <ul style="list-style-type: none"> <li>– Pollen</li> <li>– Baumringe usw.</li> </ul>			

Figur 1: Typologie klimageschichtlicher Daten



Figur 2a: Programm-Ablauf, Teil I



Figur 2b: Programm-Ablauf, Teil II

## Zeit/Ort/Quelle

Feld	Kolonne		
1	01-03	Jahr:	Hunderter, Zehner, Einer (Beispiel: Jahr 1443 ⇒ 443)
2	04-05	Monat:	01 – 012 (Januar – Dezember)
		Jahreszeiten:	13: Winter 14: Frühjahr 15: Sommer 16: Herbst 17: Jahreszeit unbestimmt
3	06	Dekade:	1 – 3 (bei Monatswerten und Jahreszeiten: 0)
4	08-09	Region:	s. Karte auf separatem Blatt
5	11-14	Quelle:	entsprechend der Bibliographie (= Leitnummer auf den entsprechenden Blättern)

## Tägliche Beobachtungen der Witterung

je drei Felder für Luftmasse, Thermik und Witterungscharakter

6	16-18	1. Tag	⇒ s. Spezialcode
7	20-22	2. Tag	⇒ s. Spezialcode
8	24-26	3. Tag	⇒ s. Spezialcode
9	28-30	4. Tag	⇒ s. Spezialcode
10	32-34	5. Tag	⇒ s. Spezialcode
11	36-38	6. Tag	⇒ s. Spezialcode
12	40-42	7. Tag	⇒ s. Spezialcode
13	44-46	8. Tag	⇒ s. Spezialcode
14	48-50	9. Tag	⇒ s. Spezialcode
15	52-54	10. Tag	⇒ s. Spezialcode

## Interpretationsstufen von Quellen

16	56	7: großzügige Interpretation (sehr unsicher)	im Output ⇒ **
		8: wahrscheinlich	im Output ⇒ *
		-: Quelle sicher (zeitgenössisch)	im Output ⇒ ohne Zeichen, Beobachter und Quelle

Figur 3a: Codebuch-Übersicht, Teil I



## Charakterisierung von Dekaden, Monaten oder Jahreszeiten und von Phänomenen, die dekadewise codiert werden

17	57-58	dominante Winde	⇒ s. Spezialcode
18	59-60	dominante Thermik	⇒ s. Spezialcode
19	62-63	Witterungscharakter	⇒ s. Spezialcode
20	64	Schneedecke	
21	66-67	besondere Naturereignisse	
22	68	Tagescode für Phänologie	
23	69-70	Phänologische Beobachtungen	
24	71-72	Witterungsbedingte Schäden an Gebäuden, Vieh und Kulturen	
25	73-74	Grösse der Ernten, Lebensmittelpreise, Epidemien	

## Kalender, Kommentare, Verweise, Varia

26	75	1: Gregorianischer Kalender mit Kommentar
		2: Julianischer Kalender mit Kommentar

Figur 3b: Codebuch-Übersicht, Teil II

## Literatur

*Braudel, F.*: La Méditerranée et le monde méditerranéen au temps de Philippe II, Paris, 1949.

*Mols, R.*: Die Bevölkerung Europas 1500–1700, In: Europäische Wirtschaftsgeschichte (= The Fontana Economic History of Europe, Bd. 2: Sechzehntes und siebzehntes Jahrhundert.) Stuttgart, 1979, S. 5-49.

*Pfister, Ch.*: Das Klima der Schweiz von 1525 bis 1860 und seine Bedeutung in der Geschichte von Bevölkerung und Landwirtschaft, Band 1: Klimageschichte der Schweiz, Band 2: Bevölkerung, Klima und Agrarmodernisierung, Bern: Paul Haupt, 1984.

*Pfister, Ch.*: Klimageschichte Datenbank CLIMHIST, Bd. 1: Schweiz 1525-1863. Bern, 1984 (Zu beziehen bei: METEOTEST, Hallerstr. 12, CH-3012 Bern).

1684 J A N U A R

1. DEKADE

KUEHL. VORM. SCHNEE (VEREINZELT REGEN). R: 6 ZUERICH:

408 M(FRIES,Q 111),J.

EINGESCHNEIT. R: 6 ZUERICH: 408 M(FRIES,Q 111),J.R:10,

EINSIEDELN: 882 M(DIETRICH,Q 52)\*\*

2. DEKADE

KALT/UNBEKANNT. R:10 EINSIEDELN: 882 M(DIETRICH,Q 52)\*\*

SEHR KALT. VORM.SONNIG(KUERZERE SCHNEEPERIODE). R: 6 ZUERICH:

408 M(FRIES, Q 111),J.

DAUERENDE SCHNEEDECKE. R: 6 ZUERICH: 408 M(FRIES, Q 111),J.

ANZEICHEN VON STARKER KAELE ZLLE.GEWAESSER ZUGEFROREN UND BEGEBAR

R: 6 ZUERICH: 408 M(FRIES,Q 111),J.R.: 10 EINSIEDELN:

882 M(DIETRICH,Q 52)\*\*

3. DEKADE

KUEHL. VORM.BEWOLKT(KUERZERE SCHNEEPERIODE). R: 6 ZUERICH:

408 M(FRIES, Q 111),J.

DAUERENDE SCHNEEDECKE. R: 6 ZUERICH: 408 M(FRIES, Q 111),J.

GESAMTER MONAT

SEHR KALT. R:13 RM MAIENFELD(Q 49),J.R: 2,ST.GALLEN:

670 M(GALLUS II.,Q 56) WINTERTHUR: 442 M(BUCHER,Q 81), J.(A272),

WINTERTHUR: 442 M(Q 84),J.(A273)R: 8,RM. NEUCHATEL(BOYVE,Q 191), J.

DAUERENDE SCHNEEDECKE. R: 2 ST.GALLEN: 670 M(GALLUS II.,Q 56)

ANZEICHEN VON STARKER KAELE ALLE GEWAESSER ZUGEFROREN UND BEGEBAR.

R: 13 RM MAIENFELD(Q 49),J.R.: 1,BASEL: 269 M(Q 3),J.(A274)

R: 2, WINTERTHUR: 442 M(BUCHER,Q 81),J. R: 8,RM NEUCHATEL

(BOYVE,Q 191),J.

FROSTSCHADEN REBEN. R: 2 WINTERTHUR: 442 M(Q 84),J.(A273)

FROSTSCHADEN . R:13 RM MAIENFELD(Q 49),J.

TROCKEN 9 N-TAGE(M:14TG) VORWIEGEND SCHNEE:88% KALT

SEHR LANGE SCHNEEBEDECKUNG 27 TG.

R: 6 ZUERICH: 408 M(FRIES,Q 111),J.#

T-INDEX: -3, SEHR KALT . N-INDEX: -1, UNTERNORMAL

Figur 4: Deutscher CLIMHIST-Output

1557 F E V R I E R

1ERE PERIODE DE 10 JOURS

VARIABLE(JOURS ISOLES DE NEIGE OU NEIGE ET PLUIE). R: 6 ZUERICH:

430 M(W HALLER, S 174)

EPAISSE COUCHE DE NEIGE PLUSIERS JOURNEES . R: 6 ZUERICH:

430 M(W HALLER, S 174)

NEIGE/DEGEL SOUDAIN/INONDATION. R: 2 WINTERTHUR: 442 M(U MEYER S 89)

2EME PERIODE DE 10 JOURS

PAS DE RELEVÉ(JOURS ISOLES DE NEIGE OU NEIGE ET PLUIE). R 6 ZUERICH:

430 M(W HALLER, S 174)

ENNEIGE. R: 6 ZUERICH: 430 M(W HALLER, S 174)

3EME PERIODE DE 10 JOURS

FRAIS/PAS DE RELEVÉ. PAS DE RELEVÉ(JOURS ISOLES DE NEIGE OU

NEIGE ET PLUIE). R 6 ZUERICH: 430 M(W HALLER, S 174)

EPAISSE COUCHE DE NEIGE PLUSIEURS JOURNEES . R: 6 ZUERICH:

430 M(W HALLER, S 174)

LE MOIS EN TOUT

VARIABLE 12 JOURS DE FR(M:10JOURS)R: 6 ZUERICH: 430 M(W HALLER, S 174)#

INDICE DES TEMP: 0, NORMAL . INDICE DES PREC: 0, NORMAL

Figur 5: Französischer CLIMHIST-Output

**Historia Vaga**

**Ein computergestütztes Projekt zur Migrationsgeschichte  
des 15. und 16. Jahrhunderts**

**1. Vorbemerkung**

Es ist obsolet geworden, festzustellen, daß der Computer zu einem anerkannten Instrument der Historischen Forschung geworden ist. Selbst in Ländern wie Österreich, das in Bezug auf internationale Tendenzen und Entwicklungen in der Geschichtswissenschaft zumeist im Zustand einer Marginalität verharret, findet er bereits gelegentlich Verwendung in geschichtswissenschaftlichen Projekten<sup>1</sup>. Im Zusammenhang mit diesem Selbstverständlich-Werden von Computeranwendung in der internationalen historischen Forschung – sowohl in ihrer numerischen als auch in ihrer nichtnumerischen Ausprägung – ergab sich nach der Phase des “Kampfes” um die Durchsetzbarkeit innerhalb der Disziplinen und der Phase der Bewältigung technischer Probleme eine Tendenz zur Rückkehr zur theorielosen und nicht problemorientierten Geschichtswissenschaft<sup>2</sup>, gegen die viele Historiker im deutschsprachigen Raum in den 60er Jahren revoltiert hatten<sup>3</sup>. Aus diesen Tendenzen entstanden mitunter jene “Datenfriedhöfe”, vor denen der Veranstalter dieser Tagung gelegentlich gewarnt hat<sup>4</sup>. Die Möglichkeit der Verarbeitung großer

---

<sup>1</sup> Vgl. zuletzt: Bericht über den sechzehnten Österreichischen Historikertag in Krems/Donau (=Veröffentlichungen des Verbandes Österreichischer Geschichtsvereine 25), o.O., 1985, 599-714.

<sup>2</sup> Vgl. zur generellen Entwicklung A.G. Bogue: Numerical and Formal Analysis in the United States History, in: Journal of Interdisciplinary History 12 (1981), 137 – 175 und D. Herlihy: Numerical and Formal Analysis in European History, in: ebda. 115 – 136; zur von uns angesprochenen Tendenz vgl. z.B. K.M. Schmidt: Errungenschaften, Holzwege und Zukunftsmusik auf dem Gebiet der inhaltlichen Textanalyse mit Hilfe des Elektrorechners, in: P. Sappeler u. E. Strabner (Hrsg.): Maschinelle Verarbeitung altdeutscher Texte III, Tübingen, 1980, 101 – 111.

<sup>3</sup> Vgl. zusammenfassend J. Rüsen: Theory of History in the Development of West German Historical Studies: A Reconstruction and Outlook, in: German Studies Review 7 (1984), 11 – 25.

<sup>4</sup> Z.B. M. Thaller: Mittelalterliche Realienkunde und EDV, in: Die Erforschung von Alltag und Sachkultur des Mittelalters. Methode – Ziel – Verwirklichung. Internationales Round-Table-Gespräch Krems an der Donau 20. September 1982 (=Veröff. d. Inst. f. ma. Realienkunde 6 = Sb. Ak. Wien, phil.-hist. Kl.

Daten- bzw. Quellenmengen durch den Rechner scheint auf viele Historiker genau die Art von Faszination auszuüben, die früher jene gewaltigen Mengen von Zettelkästen hatten, deren Verwaltung und Benützung so aufwendig war, daß für die eigentliche geschichtswissenschaftliche Arbeit wenig Zeit blieb. Wenn hier ein Projekt vorgestellt werden soll, dessen Eigenschaft es ebenfalls ist, einige -zig Tausend von Einzelquellen maschinenlesbar zu machen, muß erläutert werden, wie weit Computer, EDV und quantifizierende Techniken nicht zum reinen Selbstzweck zu werden drohen.

Im folgenden sollen hier also nicht technische Details unter dem Aspekt der Datenverarbeitung behandelt, sondern versucht werden, unser Projekt, das sich mit Migrationsgeschichte im österreichischen Raum im Spätmittelalter und in der frühen Neuzeit befaßt, auch bezüglich wesentlicher Punkte der Gesamtkonzeption und deren praktischer Umsetzung darzustellen. Daraufhin soll gezeigt werden, in welcher Weise dieses Projekt das Problem der formalen Beschreibung von Quellen löst<sup>5</sup>, und in welcher Form damit auch Quellen für Sekundäranalysen und andere Forschungsprojekte zur Verfügung gestellt werden können.

## **2. Voraussetzungen:**

### **2.1. Möglichkeiten interdisziplinären Arbeitens: das Angebot der Sozialwissenschaft.**

In der sozialwissenschaftlichen Forschung ist das Studium von Migrationsprozessen und geographischer Mobilität seit langem ein vielbearbeitetes Thema<sup>6</sup>. Die Deskription und Prognose von Wanderungsphänomenen ist

---

433), Wien, 1984, 219 – 225, hier 223f.; – *ders.*: DESCRIPTOR. Probleme der Entwicklung eines Programmsystems zur computerunterstützten Auswertung mittelalterlicher Bildquellen, in: Europäische Sachkultur des Mittelalters (=Veröff. d. Inst. f. ma. Realienkunde 4 = Sb. Ak. Wien, phil.-hist. Kl. 374), Wien, 1980, 167 – 194, hier 172f.

<sup>5</sup> *Manfred Thaller* sind wir für die aktive Unterstützung und geduldige Beratung zu großem Dank verpflichtet.

<sup>6</sup> Vgl. die Bibliographien in *P. Franz*: Soziologie der räumlichen Mobilität. Eine Einführung, Frankfurt / New York, 1984; – *J. Bähr*: Bevölkerungsgeographie, Stuttgart, 1983; – *W. Kuls* (Hrsg.): Probleme der Bevölkerungsgeographie (=Wege der Forschung 468), Darmstadt, 1978; – *J. Leib* u. *G. Mertins*: Bevölkerungsgeographie (=Das Geographische Seminar), Braunschweig, 1983; – *G. Albrecht*: Soziologie der geographischen Mobilität, Stuttgart, 1972; – *H. Esser*: Aspekte der Wanderungssoziologie. Assimilation und Integration von Wanderern. Eine hand-

häufig Teil jener Forschungsaufträge an die Sozialwissenschaft, deren Ergebnisse – vor allem, wenn es um das Problem der sozialen und politischen Integration geht – eine für politische Entscheidungen wichtige Rolle einnehmen können<sup>7</sup>.

Deutlich – so erscheint es dem im Nachbargebiet bibliographierenden Historiker – ist auch die quantitative Überlegenheit empirischer Forschungen und regional sowie zeitlich häufig eng begrenzter Datenstudien gegenüber generellen, Modelle und Theorien bildenden Arbeiten<sup>8</sup>.

Um in gesellschaftspolitischer Planung Aufgabenbereiche wahrnehmen zu können, wurde von der sozialwissenschaftlichen Migrationsforschung eine Fülle stets komplexer werdender Maßzahlen entwickelt, die Migrationsprozesse auf verschiedenen Niveaus beschreibbar gemacht hat<sup>9</sup>. Diese Maßzahlen beruhen fast immer auf den Voraussetzungen guter und kontinuierlicher Daten, wie sie frühestens durch das Census- und Meldewesen des 19. Jahrhunderts gegeben sind.

Das schränkt die unmittelbare Übernahme sozialwissenschaftlicher Konzepte für unser Projekt erheblich ein, da die im Projekt verwendeten, wie viele in der Mittelalterforschung zur Verfügung stehenden Quellen keine hin-

---

lungstheoretische Analyse, Darmstadt / Neuwied, 1980; – P. Weber: Geographische Mobilitätsforschung (=Erträge der Forschung 179), Darmstadt, 1982; – K. Horstmann: Zur Soziologie der Wanderungen, in: R. König: Handbuch der empirischen Sozialforschung, Bd. 5 der Taschenbuchausgabe, 2. Aufl., München, 1976, 104 – 186.

<sup>7</sup> Zu den Methoden der Prognose vgl. z.B. T.W. Rogers: Migration Prediction on the Basis of Prior Migratory Behavior: A Methodological Note, in: International Migration 7 (1969), 13 – 19.

<sup>8</sup> Wie Anm. 6. Vgl. den Überblick in den beiden Themenheften: Migration in Europe 1 and 2. Trends in Research and Sociological Approaches. Perspectives from the Countries of Origin and Destination (1960-1983), in: Current Sociology 32, 2 (1984) und 32, 3 (1984).

<sup>9</sup> Vgl. z.B. Y. Tugault: La mesure de la mobilité. Cinq études sur les migrations internes (=Institut National d'Etudes Demographiques. Travaux et Documents. Cahiers N. 67), Paris, 1973; – H.-J. Hoffmann-Novotny: Migration. Ein Beitrag zu einer soziologischen Erklärung, Stuttgart, 1970, 64 – 72 und ausführlich W. Schweitzer: Modelle zur Erfassung von Wanderungsbewegungen (=Schriften zur wirtschaftswissenschaftlichen Forschung 129), Meisenheim am Glan, 1978. Für die Geschichtswissenschaft vgl. z.B. J.A. Agnew u. K.R. Cox: Urban In-Migration in Historical Perspective. An Approach to Measurement, in: Historical Methods 12 (1979), 145 – 155.

reichende Möglichkeit zur operationalisierten Umsetzung als quantitative Daten bieten. Nichtsdestoweniger müssen hier aber die Arbeiten von *D. Kubat* und *H.J. Hoffmann-Novotny* Erwähnung finden, deren Ansätze so komplex sind, daß sie – ungeachtet des Problems der empirischen Umsetzbarkeit dieser Ansätze – die Ausgangsüberlegungen des Projektes erheblich mitbeeinflußt haben<sup>10</sup>.

## 2.2. Bemerkungen zur Forschungsgeschichte in den Historischen Wissenschaften

Wie in den Sozialwissenschaften ist auch von Historikern die Wichtigkeit der Geschichte der Wanderungen anerkannt. *W. Köllmann* spricht in diesem Zusammenhang von "geschichtsmächtigen Phänomenen" universeller Bedeutung<sup>11</sup>. Ohne daß hier auch nur andeutungsweise ein Forschungsbericht gegeben werden soll, muß festgestellt werden, daß das Interesse für Migrationsgeschichte vor allem auf zwei Komplexe beschränkt gewesen ist: einerseits auf den umfassenden Bereich der transatlantischen Migration, der in vielen Fällen mit genealogischen Intentionen jener Art verbunden ist, die sich im Ausforschen der biologischen Vorfahren auf der Suche nach der Heimat in der Geschichte befindet; andererseits auf den – im Zusammenhang mit modernisierungstheoretischen Integrationen stehenden – Phänomenen der Land – Stadt – Migration während des Industrialisierungsprozesses.

Die Beschäftigung mit globalen Wanderungsbewegungen des Mittelalters, etwa der deutschen "Ostbewegung" war häufig getragen von einem politisch motivierten Legitimationsinteresse, das territoriale Ansprüche und eine expansionistische Politik rechtfertigen sollte<sup>12</sup>. Bis zu einem gewissen Grad ist die Migrationsgeschichte des Mittelalters aus diesem Grund diskreditiert.

---

<sup>10</sup> *D. Kubat u. H.-J. Hoffmann-Novotny*: Migration: towards a new Paradigm, in: *International Social Science Journal* 33 (1981), 307 – 329 und *dies.*: Migrations: vers un nouveau paradigme, in: *Revue internationale Science sociologique* 33 (1981), 335 – 359.

<sup>11</sup> *W. Köllmann*: Versuch eines Entwurfs einer historisch-soziologischen Wanderungstheorie, In: *U. Engelhardt, V. Sellin u. H. Stuke* (Hrsg.): *Soziale Bewegung und politische Verfassung. Beiträge zur Geschichte der modernen Welt (=Industrielle Welt, Sonderbd.)*, Stuttgart, 1973, 260 – 269, hier: 260.

<sup>12</sup> Vgl. anschaulich die Literatur bei *H. Grundmann*: Wahlkönigtum, Territorialpolitik und Ostbewegung im 13. und 14. Jahrhundert 1198–1378, In: *Gebhard's Handbuch der deutschen Geschichte*, Band 5 der Taschenbuchausgabe, München, 1973, 260.

### 3. Zur Konzeption des Projekts im Rahmen "neuer Ansätze" in der Geschichtswissenschaft

Hier soll nun versucht werden, das Problem der Migrationsgeschichte des Spätmittelalters und der frühen Neuzeit im Konzept zweier neuerer Forschungsansätze zu sehen; dem der "Geschichte als Historischer Sozialwissenschaft" und dem der Neuen Alltagsgeschichte. Diese beiden Ansätze wurden bisweilen als fast zwangsläufig kontroversiell angesehen<sup>13</sup>. Die Verfasser geraten bei dem Versuch, zwischen beiden Forschungsansätzen vermitteln zu wollen, mitunter in die Angriffslinie beider. Nun sind diese beiden Begriffe nicht etwa klar definiert oder auch nur definierbar<sup>14</sup>. Noch dazu wurden – zumindest im deutschsprachigen Raum – diese Ansätze fast ausschließlich von Historikern vertreten die sich mit gesellschaftlichen Phänomenen im, 19. und 20. Jahrhundert befassen<sup>15</sup>, während deutschsprachige Mediävisten und Frühneuzeitler sich bisher zum größten Teil eines Beitrages zur Diskussion enthalten haben<sup>16</sup>.

Migrationsgeschichte als ein zumindest teilweise von den Sozialwissenschaften abgeleitetes Thema legt eine Konzeptualisierung innerhalb der Historischen Sozialwissenschaft nahe. Die Historische Sozialwissenschaft, jene –

---

<sup>13</sup> Zur Kontroverse im anglophonen Raum vgl. v.a. L. Stone: *The Revival of Narrative: Reflections on a New Old History*, in: *Past and Present* 85 (1979), 3 – 24; – E. Hobsbawm: *The Revival of Narrative: Some Comments*, in: *Past and Present* 86 (1980), 3 – 8; – P. Abrams: *History, Sociology, Historical Sociology*, in: *Past and Present* 87 (1980), 3 – 16, – J.M. Kousser: *the Revivalism of Narrative: A Response to Recent Criticisms of Quantitative History*, in: *Social Science History* 8 (1984), 133 – 150. – Zur Diskussion im deutschsprachigen Raum vgl. v.a. H. Süssmuth (Hrsg.): *Historische Anthropologie*, Göttingen, 1984 und die Beiträge in *Geschichte und Gesellschaft* 10 (1984), H. 3.

<sup>14</sup> Vgl. unten, sowie für die Historische Sozialwissenschaft G. Iggers: *Neue Geschichtswissenschaft. Vom Historismus zur Historischen Sozialwissenschaft. Ein internationaler Vergleich*, München, 1978, 134f.

<sup>15</sup> Für die Historische Sozialwissenschaft vgl. für den deutschsprachigen Raum die von W. Bick, P.J. Müller und H. Reincke unter dem Titel "Quantum. Dokumentation" bzw. "Historische Sozialforschung – Historical Social Research. Dokumentation" in der Reihe *Historisch Sozialwissenschaftliche Forschungen* seit, 1977 herausgegebenen Projektdokumentationen. Für die Alltagsgeschichte siehe die Literatur bei H. Medick: "Missionare im Ruderboot?" *Ethnologische Erkenntnisweisen als Herausforderung an die Sozialgeschichte*, in: *Geschichte und Gesellschaft* 10 (1984), 295 – 319.

<sup>16</sup> Vgl. Anm. 23 und 34.



fast möchte man sagen: typisch deutsche – Reinterpretation der von vielen Historikern gefürchteten “social science history” auf der Basis der gezielten Rezeption der deutschsprachigen Klassiker der Sozialwissenschaften<sup>17</sup>, hat sich im deutschsprachigen Raum als einzige namhafte neuere historische “Schule” zu den Methoden und Techniken der Quantifizierung und der Computeranwendung bekannt<sup>18</sup>. “Historische Sozialwissenschaft (...) untersucht die zeitlichen Veränderungen menschlichen Seins und gesellschaftlicher Verhältnisse auf der Grundlage von systematisch-sozialwissenschaftlichen Problemstellungen, Methoden und Theorien.”<sup>19</sup> Diese vorsichtige Definition von G. Botz steht mit den meisten anderen Definitionsversuchen im Einklang<sup>20</sup>.

Unter diesen Voraussetzungen realisierte die Historische Sozialwissenschaft sich selbst vor allem dort, wo es um die systematische Kritik am traditionellen Historismus, und um die theoretische und empirische Auseinandersetzung mit schicht- und klassenorientierten Gesellschaftsmodellen, mit modernisierungstheoretischen Konzepten vor allem im Zusammenhang mit der Geschichte der Industrialisierung, mit Konzepten sozialen Wandels, mit gesellschaftlichen “Überbau“-Phänomenen und vor allem mit sozialen Ursachen politischer Prozesse ging<sup>21</sup>.

<sup>17</sup> Vgl. Iggers: *Neue Geschichtswissenschaft*, 130ff.

<sup>18</sup> Z.B. J. Kocka: Quantifizierung in den Geschichtswissenschaften, in: H. Best u. R. Mann (Hrsg.): *Quantitative Methoden in der historisch-sozialwissenschaftlichen Forschung* (= Historisch Sozialwissenschaftliche Forschungen 3), Stuttgart, 1977, 4 – 10 und ders.: Theories and Quantification in History, in: *Social Science History* 8 (1984), 169 – 178.

<sup>19</sup> G. Botz: Vorwort zu M. Thaller: *Numerische Datenverarbeitung für Historiker* (=Materialien zur Historischen Sozialwissenschaft 1), Wien Köln, 1982, I – X, hier VI.

<sup>20</sup> H.-U. Wehler: *Geschichte als Historische Sozialwissenschaft*, Frankfurt, 1973, 9 – 44, – R. Rürup, Zur Einführung, in: ders. (Hrsg.): *Historische Sozialwissenschaft. Beiträge zur Einführung in die Forschungspraxis*, Göttingen, 1977, 5 – 15; – J. Kocka: *Historisch-anthropologische Fragestellungen – ein Defizit der Historischen Sozialwissenschaft?*, in: H. Süßmuth: *Historische Anthropologie*, Göttingen, 1984, 73 – 83. – Vgl. auch D. Ruloff: *Geschichtsforschung und Sozialwissenschaft. Untersuchung zur Wissenschafts- und Forschungskonzeption in Historie und Politikologie*, München, 1984, 260 – 277.

<sup>21</sup> Vgl. die Jahrgänge der Zeitschriften *Geschichte und Gesellschaft* seit 1975, *Historical Social Research* seit 1977 sowie die Reihe *Historisch Sozialwissenschaftliche Forschungen* seit 1977. – Zu einer historisch-sozialwissenschaftlichen Migrationsgeschichte vgl. z.B. D. Langewiesche: *Wanderungsbewegungen in der Hochindustrial-*

Dies ist nicht der Ort, eine umfassende Würdigung der Historischen Sozialwissenschaft zu leisten. Explizit festgehalten sei ihr Verdienst um eine "Verwissenschaftlichung" der Geschichtsforschung gegenüber dem Historismus durch das Bemühen um eine reflektierte Begrifflichkeit, die (im Gegensatz etwa zu *O. Brunner*<sup>22</sup>) Trennung von Metasprache und Objektsprache und den (teilweise) sorgsamem Umgang im Messen des empirischen Gehalts von Theorien unter Verwendung quantifizierender Methoden.

Innerhalb der deutschsprachigen Mittelalter- und Frühneuzeitforschung gab es kaum eine Auseinandersetzung über die Konzepte der Historischen Sozialwissenschaft<sup>23</sup>. Das Innovationspotential des sozialgeschichtlich orientierten Teils der deutschsprachigen Mediävistik beschränkte sich – wenn überhaupt – auf die Rezeption des verschwommenen Strukturbegriffes *O. Brunners*<sup>24</sup> oder jenes stark vereinfachten Schichtbegriffes *E. Maschkes*<sup>25</sup>,

---

lisierungsperiode. Regionale, interstädtische und innerstädtische Mobilität 1880 – 1914, in: VSWG 64 (1977) 1 – 40; – *J. Brockstedt*: Regionalgeschichtliche Forschungsansätze zur geographischen Mobilität in Schleswig-Holstein, in: *Historical Social Research* 14 (1980), 34 – 37; – *ders.*: Regionale Mobilität, Wirtschaftsentwicklung und Sozialstruktur in Schleswig-Holstein (1880 – 1864), in: *W.H. Schröder* (Hrsg.): *Moderne Stadtgeschichte* (=Historisch Sozialwissenschaftliche Forschungen 8), Stuttgart, 1979; – *ders.* (Hrsg.): *Regionale Mobilität in Schleswig-Holstein 1600 – 1900. Theorie, Fallstudien, Quellenkunde, Bibliographie* (=Studien zur Wirtschafts- und Sozialgeschichte Schleswig-Holsteins 1), Neumünster, 1979.

<sup>22</sup> *O. Brunner*: *Land und Herrschaft. Grundfragen der territorialen Verfassungsgeschichte Österreichs im J. Meran: Theorien in der Geschichtswissenschaft*, Göttingen, 1985, 111ff. Nicht diese Mittelalter, 4. Aufl., Wien / Wiesbaden, 1959, 440. Ansicht teilt *O.G. Oexle*: *Sozialgeschichte – Begriffsgeschichte – Wissenschaftsgeschichte. Anmerkungen zum Werk O. Brunners*, in: VSWG 71 (1984), 305 – 341.

<sup>23</sup> Als Ausnahme vgl. etwa *F. Irsigler* (Hrsg.): *Quantitative Methoden in der Wirtschafts- und Sozialgeschichte der Vorneuzeit* (=Historisch Sozialwissenschaftliche Forschungen 4), Stuttgart, 1978 oder etwa die unter Anm. 26 genannte Literatur.

<sup>24</sup> *O. Brunner*: *Das Problem einer europäischen Sozialgeschichte*, in: *ders.*: *Neue Wege der Sozialgeschichte. Vorträge und Aufsätze*, Göttingen, 1956, 7 – 32, hier: 9.

<sup>25</sup> *Maschkes* wesentliche Arbeiten wurden wieder abgedruckt in *E. Maschke*: *Städte und Menschen. Beiträge zur Geschichte der Stadt, der Wirtschaft und Gesellschaft, 1959-1977* (=VSWG-Beiheft 68), Wiesbaden, 1980, 157-169, 275-305 und 306-379.

der erst durch eine von M. Mitterauer, 1977 ausgelöste Diskussion verfeinert werden konnte<sup>26</sup>.

Unter diesem Dilemma (fast) der gesamten Disziplin hat natürlich auch dieses Forschungsprojekt zu "leiden". Für das Problem der Migrationsgeschichte des Mittelalters und der Frühneuzeit scheint es unerlässlich, mit Begriffen des sozialen Wandels, der sozialen Schichtung, der Klassenstruktur und der Modernisierung im Sinn von Arbeitshypothesen umzugehen, auch wenn sie von der Forschung nicht hinreichend für den in Rede stehenden Zeitraum empirisch umgesetzt worden sind.

Migration als Kategorie der Alltagsgeschichte fand in der bisherigen Forschung nur marginales Interesse bzw. überhaupt keine Berücksichtigung<sup>27</sup>.

---

<sup>26</sup> M. Mitterauer: Probleme der Stratifikation in mittelalterlichen Gesellschaftssystemen, in: J. Kocka (Hrsg.): Theorien in der Praxis des Historikers. Forschungsbeispiele und ihre Diskussion (=Geschichte und Gesellschaft, Sonderh. 3), Göttingen, 1977, 13 - 43; - vgl. dazu H. Wunder: Probleme der Stratifikation in mittelalterlichen Gesellschaftssystemen. Ein Diskussionsbeitrag zu Thesen von M. Mitterauer: in: Geschichte und Gesellschaft 4 (1978), 542 - 550; - J. Ellermeyer: "Schichtung" und "Sozialstruktur" in spätmittelalterlichen Städten. Zur Verwendung sozialwissenschaftlicher Kategorien in historischer Forschung, in: Geschichte und Gesellschaft 6 (1980), 125 - 149; - E. Weihrauch: Über soziale Schichten, in: I. Batori (Hrsg.): Städtische Gesellschaft und Reformation. Kleine Schriften 2 (=Spätmittelalter und Frühe Neuzeit 12), Stuttgart, 1980, 5 - 57.

<sup>27</sup> Als Beispiel für die wenigen umfassenderen Arbeiten vgl. C. Küther: Menschen auf der Straße. Vagierende Unterschichten in Bayern, Franken und Schwaben in der zweiten Hälfte des 18. Jahrhunderts, Göttingen, 1983; - H. Hundsbichler: Realien zum Thema "Reisen" in den Reisetagebüchern des Paolo Santonino (1485-1487), in: Die Funktion der schriftlichen Quelle in der Sachkulturforschung (=Veröff. d. Inst. f. ma. Realienkunde 1 = Sb. Ak. Wien, phil.-hist. Kl. 304/4), Wien, 1976, 55 - 143; - ders.: Reise, Gastlichkeit und Nahrung im Spiegel der Reisetagebücher des Paolo Santonino (1485-1487), Phil. Diss., Wien, 1979; - H. Kühnel, Die adelige Kavaliertour im 17. Jahrhundert, in: Jahrb. f. Landeskunde v. Niederösterreich 36 (1964), 364 - 384; - G. Stretton: The Travelling Household in the Middle Ages, in: The Journal of the British Archeological Association, N.S. 40 (1935), 75 - 103. Vgl. etwa auch die im Rahmen einer "History of Civilization" präsentierte Artikelsammlung A. P. Newton (ed.): Travel and Travellers of the Middle Ages, Ndr. der 3. Aufl., 1949, London, 1968; - Peter Moraw (Hrsg.): Unterwegs sein im Mittelalter (=Zeitschrift für Historische Forschung, Beih. 1), Berlin, 1985. Als Beispiele für entsprechende Abschnitte in umfangreicheren "Alltagsgeschichten" vgl. Ch.-E. Dufourcq: La vie quotidienne dans les ports

Ausschlaggebend dafür erscheinen unter anderem sicherlich auch die häufig fehlenden Konzepte einer Alltagsgeschichte, welche gerade erst in jüngster Zeit massiv zur Diskussion gelangen. Betrachtet man diese oft vehement geführte Diskussion zur (Neuen) Alltagsgeschichte, sieht man sich einer mitunter unabsehbar scheinenden Anzahl von Alleinvertretungsansprüchen und Vereinnahmungstendenzen konfrontiert<sup>28</sup>. Die schwere Faßbarkeit bzw. Undefinierbarkeit des Alltagsbegriffs tut dazu das ihre<sup>29</sup>. Darüber hinaus wird die "Alltagsgeschichte" und die darüber entbrannte Diskussion beinahe stillschweigend mit dem Zeitraum des 19. und 20. Jahrhunderts verbunden<sup>30</sup>. Allerdings gewinnt auch der Alltag in der Antike oder im Mittelalter ver-

---

méditerranéens au Moyen Age, Paris, 1975, 83 - 98; - H. Kühnel (Hrsg.): Alltag im Spätmittelalter, 2. Aufl., Graz / Wien / Köln, 1985, 114 - 120. In einer der Kultur- und Sittengeschichte verbundenen Tradition deutscher Geschichtsschreibung ab der zweiten Hälfte des 19. Jahrhunderts zeigen sich die Schwerpunkte zu einer Untersuchung von migrationsgeschichtlichen Aspekten in der Edition und Bearbeitung unterschiedlichster Reiseberichte (vgl. die bibliographischen Angaben bei *Hundsichler: Reise XI-XLIV*).

<sup>28</sup> Vgl. die jüngsten Überblicke zur Diskussion bei K. Tenfelde: Schwierigkeiten mit dem Alltag, in: *Geschichte und Gesellschaft* 10 (1984), 376 - 394; - J. Kocka: Zurück zur Erzählung? Plädoyer für historische Argumentation, in: ebda, 395 - 408; - H. Süßmuth (Hrsg.): *Historische Anthropologie*, Göttingen, 1984; - ferner die Beiträge von D.J.K. Peukert, L. Nagl, R. Sieder und R. Wirtz, in: H. Nagl-Docekal u. F. Wimmer (Hrsg.): *Neue Ansätze in der Geschichtswissenschaft (=conceptus-studien 1)*, Wien, 1984, und - reichlich oberflächlich und teilweise sachlich falsch - H.Ch. Ehalt: *Geschichte von unten*, in: ders. (Hrsg.): *Geschichte von unten. Fragestellungen, Methoden und Projekte einer Geschichte des Alltags (=Kulturstudien 1)*, Wien / Köln / Graz, 1984, 11-39.

<sup>29</sup> Vgl. Tenfelde, *Schwierigkeiten*, 385 - 391, sowie die kurzen Bemerkungen von N. Elias, auf die sich Definitions-, Eingrenzungs- und Beschreibungsversuche des historischen Alltags noch immer regelmäßig stützen: N. Elias: Zum Begriff des Alltags, in: K. Hammerich - M. Klein (Hrsg.): *Materialien zur Soziologie des Alltags (=Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderh. 20)*, Opladen, 1978, 22 - 29. Vgl. auch Anm. 32 und 36.

<sup>30</sup> Dies heißt jedoch nicht, daß auf gewisse herausragende Arbeiten, die sich mit Alltag in der vorindustriellen Zeit beschäftigen, nicht regelmäßig - wenn auch häufig nur in einer kurzen Anmerkung - verwiesen wird. Dazu zählen vorrangig E. Le Roy Ladurie: *Montaillou. Ein Dorf vor dem Inquisitor*, München, 1980 und C. Ginzburg: *Der Käse und die Würmer. Die Welt eines Müllers um 1600*, Frankfurt, 1979. Vgl. eine derartige Bemerkung auch bei Tenfelde: *Schwierigkeiten*, 383.

stärkte Vermarktung<sup>31</sup>. In die Diskussion um die Alltagsgeschichte wird von dessen Vertretern jedoch kaum eingegriffen<sup>32</sup>. In manchen Fällen schließen sie recht nahtlos an die Sammelleidenschaft des ausgehenden 19. Jahrhunderts an, die sich in Monumentalwerken zum "täglichen Leben", "häuslichen Leben", "vie privée", "vie quotidienne", "daily life", etc. niederschlug<sup>33</sup>. Wandel wird nur marginal berücksichtigt. Eine Methodendiskussion findet nur in bescheidenem Rahmen statt<sup>34</sup>. Als positive Hinterlassenschaft jener "Alltagsgeschichte" des 19. Jahrhunderts, die sich in Kultur- und Sittenge-

---

<sup>31</sup> Vgl. für das Mittelalter z.B. jüngst O. Borst: *Alltagsleben im Mittelalter*, Frankfurt, 1983; – Kühnel: *Alltag* (s. Anm. 27); – W. Hansen: *Kalenderminiaturen der Stundenbücher. Mittelalterliches Leben im Jahreslauf*, München, 1984; – K.-S. Kramer: *Fränkisches Alltagsleben um 1500. Eid. Markt und Zoll im Volkacher Salbuch*, Würzburg, 1985.

<sup>32</sup> Als eklatantes Beispiel ist hierfür die Abhandlung von O. Borst (s. Anm. 31) anzusehen. Vgl. die diesbezügliche Kritik bei H.-D. Heimann: *Über Frauenbild und Frauenalltag im späten Mittelalter – oder: Vom Lob der Frau im Angesicht der Hexe*, in: *Frau und spätmittelalterlicher Alltag* (=Veröffentl. d. Inst. f. ma. Realienkunde 9 = Sb. Ak. Wien, phil.-hist. Kl.), im Druck.

<sup>33</sup> Typische Beispiele dieser Art sind: A. Schultz: *Das höfische Leben zur Zeit der Minnesinger*, Ndr. Osnabrück, 1965; – ders.: *Deutsches Leben im XIV. und XV. Jahrhundert*, Wien / Prag / Leipzig, 1892; – ders.: *das häusliche Leben der europäischen Kulturvölker vom Mittelalter bis zur zweiten Hälfte des XVIII. Jahrhunderts* (=Handbuch der mittleren und neueren Geschichte IV), Ndr. Osnabrück, 1968; – V. Gay: *Glossaire archéologique du Moyen Age et de la Renaissance*, 2 Bde., Paris, 1887/1928, – A. Franklin: *La vie privée d'autrefois. Arts et métiers. Modes, mœurs, usages des Parisiens du Xe au XVIIIe siècle d'après des documents originaux ou inédits*, 27 vol., Paris, 1887-1902; – G. Coulton: *Life in the Middle Ages*, 4 vol., 2nd ed., Cambridge, 1928. Die von K. Lambrecht geforderte, auf materielle Zustände und sozialpsychologische Aspekte aufbauende Kulturgeschichte blieb eher unzureichend. Vgl. K. Lambrecht: *Deutsches Wirtschaftsleben im Mittelalter*, 3 Bde., Ndr., Aalen, 1969. Vgl. dazu auch K.H. Metz: *Historisches "Verstehen" und Sozialpsychologie – Karl Lambrecht und seine "Wissenschaft der Geschichte"*, in: *Saeculum* 33 (1982), 95 – 104.

<sup>34</sup> Vgl. J. Le Goff: *Geschichtswissenschaft und Erforschung des Alltags*, in: ders.: *Für ein anderes Mittelalter. Zeit, Arbeit und Kultur im Europa des 5. bis 15. Jahrhunderts*, Frankfurt / Berlin / Wien, 1984, 175 – 189; – *Die Erforschung von Alltag und Sachkultur des Mittelalters. Methode – Ziel – Verwirklichung* (=Veröffentl. d. Instituts f. ma. Realienkunde 6 = Sb. Ak. Wien, phil.-hist. Kl. 433), Wien, 1984; – H.W. Goetz: *Alltag im Mittelalter. Methodische Überlegungen anlässlich einer*

schichte niederschlug bzw. aus ihr hervorging, ist der breitere Zugang anzusehen. Sie berücksichtigte zwar in entschiedenem Maße – erstmals – auch Unterschichten und Randständige, beschränkte sich jedoch keinesfalls auf sie.

Aus unserem Verständnis heraus spielen für eine Rekonstruktion historischen Alltags die folgenden zwei Kriterien eine entscheidende Rolle:

1) Der Alltag eines Menschen soll verstanden werden als Verhalten in routinisierten Bahnen. Alltagskultur ist für uns Verwirklichung und Ausführung routinisierten Verhaltens<sup>35</sup>. Sie setzt sich aus einer nicht einzugrenzenden Menge einzelner Alltäglichkeiten zusammen<sup>36</sup>. Dadurch wird offensichtlich, daß unser Bestreben dahin gehen muß, für zu untersuchende soziale Gruppen oder auch zeitliche Perioden, welcher Form auch immer, jene Alltäglichkeiten herauszufinden, welche durch ihre Mitglieder gleich oder ähnlich ausgeführt bzw. rezipiert werden. Es geht somit unter anderem darum festzustellen, in welchen Bereichen des Alltags für bestimmte Gruppen gleiche Trends bzw. gleiche oder ähnliche Ausformungen nachzuweisen sind.

Untersuchungsgegenstand von Alltagsgeschichte ist damit nach unserem Verständnis jenes Verhalten von Menschen jedweder sozialer Zugehörigkeit, gesellschaftlicher Stellung und Funktion, welches routinemäßig oder routinisiert zu Tage tritt, bzw. sind zusätzlich jene Objekte im weitesten Sinne, die im Rahmen eines solchen Verhaltens von Bedeutung sind.

2) Alltag besteht, entwickelt und verändert sich im Rahmen von bzw. durch Kommunikation<sup>37</sup>: Kommunikation in der alltäglichen Handlung, als

---

Neuerscheinung, in: Archiv f. Kulturgeschichte 67 (1985), 207-225. Vgl. auch die Bemerkungen bei Heimann: Frauenbild (s. Anm. 32).

<sup>35</sup> Vgl. Elias: Zum Begriff des Alltags 26, – P. Borscheid: Plädoyer für eine Geschichte des Alltäglichen, in: P. Borscheid u. H.J. Teuteberg (Hrsg.): Ehe, Liebe, Tod. Zum Wandel der Familie, der Geschlechts- und Generationsbeziehungen in der Neuzeit, Münster, 1983, 7f.

<sup>36</sup> Vgl. G. Jaritz: Der Einfluß der politischen Verhältnisse auf die Entwicklung der Alltagskultur im spätmittelalterlichen Österreich, in: Bericht über den 16. Österreichischen Historikertag in Krems / Donau (=Veröffentlichungen des Verbandes Österreichischer Geschichtsvereine 25), o.O., 1985, 527.

<sup>37</sup> Vgl. zusammengefaßt zu jener Relevanz des kommunikativen Elements J. Habermas: Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns, Frankfurt, 1984, bes. 594f. – Als Beispiele für die Behandlung von Kommunikation im Mittelalter, betreffend Kulturkontakt und Kulturaustausch vgl. O.J. Geanakoplos: Interaction of the "Sibling" Byzantine and Western Culture in the Middle Ages and Italian Renaissance (330 – 1600), New Haven London, 1976; – J. Richard:

Weitergabe von Wissen und Erfahrung, als Kennenlernen von Neuem, als Übernahme und Imitation, als Verbreitung, Aufnahme oder Ablehnung von auf den Alltag direkt oder indirekt bezogener Norm – kurz: Kommunikation als kultureller Austausch. Für jenes kommunikative Element sind die Komponenten von Migration und horizontaler Mobilität als besonders entscheidend anzusehen.

Wenn wir aus der Sicht der Mittelalterforschung und den in diesen Bereichen bereits vorliegenden Versuchen zu einer Rekonstruktion des Alltags von einer Neuen Alltagsgeschichte sprechen wollen, so muß diese ihre Schwerpunkte auf die bislang eher vernachlässigten Elemente von Routinisierung und Kommunikation legen und sich damit in Einklang auf die Rekonstruktion langfristiger Entwicklungsverläufe konzentrieren.

Ohne die Bedeutung des individuellen Momentes von Migration und horizontaler Mobilität – vor allem als Korrektiv – leugnen zu wollen, müssen wir uns bewußt sein, daß der Stellenwert von Migration für den Alltag und ihre Auswirkungen auf den Alltag nur dann erkannt werden können, wenn eine Untersuchung präsumptiver Massenphänomene erfolgt. Damit wird diese alltagsgeschichtliche Fragestellung zur Migration gleichzeitig zu einem mentalitätsgeschichtlichen Problem<sup>38</sup>. Darüber hinaus ergibt sich hiermit auch die Möglichkeit aufzuzeigen, inwieweit quantitative Methoden für eine dies-

---

Orient et Occident au Moyen Age: contacts et relations (XIIe – XVe s.), London, 1976; – ders.: Les relations entre l'Occident et l'Orient au Moyen Age, London, 1977; – V. Ritter: Kulturkontakte und soziales Lernen im Mittelalter, Kreuzzüge im Licht einer mittelalterlichen Biographie (= Kollektive Einstellungen und sozialer Wandel im Mittelalter 1), Köln – Wien, 1973; – F. Valgavec: Geschichte der deutschen Kulturbeziehungen in Südosteuropa I: Mittelalter (= Südosteuropäische Arbeiten 41), München, 1953; – K. Voigt: Italienische Berichte aus dem mittelalterlichen Deutschland (= Kieler Historische Studien 17), Stuttgart, 1973. Zu Sprache und Kommunikation im Mittelalter vgl. M. Richter: Sprache und Gesellschaft im Mittelalter. Untersuchungen zur mündlichen Kommunikation in England von der Mitte des 11. bis zum Beginn des 14. Jahrhunderts (= Monographien zur Geschichte des Mittelalters 18), Stuttgart, 1979; – B. Kraus: Vorstudien zur Untersuchung einfacher Formen sprachlichen Handelns im Mittelalter, phil. Diss., Karlsruhe, 1979.

<sup>38</sup> Vgl. allg. u.a. J. Le Goff: Les mentalités: une histoire ambiguë, in: ders. u. P. Nora (Hrsg.): Faire de l'histoire – Nouveaux objets, Paris, 1974, 76 – 94; – R. Reichardt, "Histoire des mentalités". Eine neue Dimension der Sozialgeschichte am Beispiel des französischen Ancien régime, in: Internationales Archiv für Sozialgeschichte der deutschen Literatur 3 (1977), 131-166.

bezügliche Untersuchung zielführend erscheinen, was von manchen Vertretern der Neuen Alltagsgeschichte entschieden bestritten wird<sup>39</sup>.

Zusammenfassend ist festzuhalten, daß in das Migrationsgeschichte-Projekt Konzepte beider Paradigmen, bewußt diese miteinander verknüpfend, eingegangen sind. Beide dieser "neuen Ansätze" generieren sinnvolle Fragen im Zusammenhang mit der Projektkonzeption, die sich aller Wahrscheinlichkeit nach aus den Quellen auch sinnvoll beantworten lassen werden.

#### 4. Zur konkreten Umsetzung im Projekt<sup>40</sup>

Migration als Massenphänomen setzt quantitative Ansätze voraus. Für den Historiker, der sich mit dem Spätmittelalter und der frühen Neuzeit befaßt, existieren "harte" statistische Daten trotz des Vorhandenseins "früher" Massenquellen nicht<sup>41</sup>. Der Ausgangspunkt quantifizierender Argumentation muß also hier aus "vielen" Einzelfällen rekonstruiert werden. Die Dokumentation und der Vergleich dieser "vielen" Einzelfälle ist eines der größten Probleme des gegenständlichen Forschungsprojekts.

Unter einem "Einzelfall" (= einem "Migrationsakt") wurde zunächst einmal die bloße Tatsache der Veränderung der "Ortsanwesenheit" einer Person oder einer (zahlenmäßig häufig nicht näher bestimmbaren) Personengruppe verstanden. Damit findet ein sehr stark extendierter Migrationsbegriff Anwendung, der den Vorteil hat, einerseits gegenüber allen Typisierungsversuchen von Migration offen zu sein<sup>42</sup> und sowohl dauerhafte als auch nicht

---

<sup>39</sup> Vgl. jüngst R. Floud: Quantitative History and People's History: Two Methods in Conflict?, in: Social Science History 8 (1984), 151 – 168.

<sup>40</sup> Seit 1984 arbeitet eine Gruppe junger österreichischer Historiker(innen) (B. Kaiser, H. Losch, I. Matschinegg, H. Mandl-Neumann, B. Rath, B. Schuh, P. Teibenbacher u. die Verf. dieses Papiers) in Zusammenarbeit mit dem Institut für mittelalterliche Realienkunde der österr. Akad. d. Wiss., Krems, und dem Ludwig-Boltzmann-Institut für Historische Sozialwissenschaft, Salzburg, und unter der Leitung von Herwig Ebner, Graz, im Rahmen eines Forschungsprojekts, das vom "Fonds zur Förderung wissenschaftlicher Forschung in Österreich" finanziert wird, an der Erforschung spätmittelalterlicher und frühneuzeitlicher Migration im österreichischen Raum (Projektnr. P 5525).

<sup>41</sup> Vgl. z.B. die Zusammenstellung bei S. Hochstadt: Migration in Preindustrial Germany, in: Central European History 16 (1983), 195 – 224.

<sup>42</sup> Vgl. z.B. R. Heberle: Zur Typologie der Wanderungen, in: W. Köllmann u. P. Marschalck (Hrsg.): Bevölkerungsgeschichte (=Neue Wissenschaftliche Bibliothek 54), Köln, 1972, 69 – 75 und die Zusammenstellung bei Hoffmann-Novotny:



dauerhafte Formen der Migration miteinschließt<sup>43</sup>, und andererseits ganz verschiedene Quellengattungen dem Projekt subsumierbar und analysierbar macht.

Derartige Migrationsakte, wie sie hier verstanden werden, lassen sich auf vielerlei Weise und in ganz unterschiedlichen Quellen, oft auch nur als singuläres, individuelles Phänomen, feststellen. Diese Einzelfälle müssen nun, um sinnvoll analysiert werden zu können, auf eine geeignete soziale Einheit, auf eine soziale Gruppe, Schicht oder Klasse, der das wandernde Individuum angehört, bezogen sein. Desgleichen bedarf es einer ständigen Evidenthaltung in Bezug auf den Quellenbestand, der der Einzelfall oder die Einzelfälle entnommen ist bzw. sind.

Arbeitet der Historiker, der sich mit Spätmittelalter und der frühen Neuzeit befaßt, an einem Archivbestand, so hat er es – wie der Soziologe – mit einer Stichprobe zu tun: nun kann hier in vielen Fällen nicht angegeben werden, wie diese Stichprobe zustande gekommen ist, auch können meist keine Angaben über seine Grundgesamtheit gemacht werden<sup>44</sup>.

So stand das Problem der Evaluierbarkeit mittelalterlichen und frühneuzeitlichen Quellenmaterials auf der Basis einer mit Bedacht durchgeführten Standardisierung unter anderem im Mittelpunkt der Überlegungen, die dem Projekt vorausgingen. Dieses Quellenmaterial kann dem "harten" Instrumentarium der empirischen Sozialwissenschaften zunächst nicht direkt zugänglich gemacht werden.

Auf die Problematik einer Projektion rezenter soziologischer Modelle auf mittelalterliche Gesellschaften hat *M. Mitterauer* unter Bezugnahme auf *O.*

---

Migration, 55 – 63. In der historischen Forschung bemüht man sich ebenfalls Typologisierungen vorzunehmen. Vgl. etwa *K.L. Reyerson*: Patterns of Population Attraction and Mobility: The Case of Montpellier, 1293 – 1348, in: *Viator* 10 (1979), 257 – 281; – *P. Mc Clure*: Patterns of Migration in the Late Middle Ages: The Evidence of English Place-Name Surnames, in: *Economic History Review* 32 (1979), 167 – 182; – *A.G. MacPherson*: Migration fields in a traditional Highland community, 1350 – 1850, in: *Journal of Historical Geography* 10 (1984), 1 – 14; – *J. Patten*: Patterns of migration and movement of labour to three pre-industrial East Anglian towns, in: *Journal of Historical Geography* 2 (1976), 111 – 129 und die Beiträge von *R. Comba*, *G. Piccinni* und *G. Pinto* (ed.): *Strutture familiari, epidemie, migrazioni nell' Italia medievale*, Napoli, 1984.

<sup>43</sup> Vgl. *Hoffmann-Novotny*: Migration, 50 – 55.

<sup>44</sup> Vgl. als Kurzfassung für Historiker, *H. Rohlinger*: Quellen als Auswahl – Auswahl der Quellen, in: *Historical Social Research* 24 (1982), 34 – 62.

Brunner nachdrücklich hingewiesen<sup>45</sup>. Im zur Rede stehenden Projekt gehen mindestens auf zwei Argumentationsebenen soziologische bzw. historisch sozialwissenschaftliche Konzepte in die Formulierungen und intendierten Lösungen sozialgeschichtlicher Problematik ein.

A. Auf der ersten Argumentationsebene werden die zu untersuchenden Personenkreise als soziale Gruppen definiert. Die Verwendung dieses Konzeptes sozialer Gruppen entspricht einerseits der Anforderung von Offenheit der Betrachtungsweise, wie sie eine mittelalterliche und frühneuzeitliche Quellen-situation verlangt<sup>46</sup>, andererseits entbehrt sie nicht jener diskriminierenden Kriterienhaftigkeit, ohne die ein sinnvolles Vorgehen nicht denkbar wäre.

Zu den einzelnen Gruppen gehören:

I. Angehörige religiöser (v.a. monastischer) Gemeinschaften

Quellen: Urkunden, Rechnungsbücher, Profefzettel-sammlungen

Das Phänomen Migration hat in dieser Gruppe vor allem drei Ausprägungen:

- a) Migration im Zusammenhang mit dem Eintritt einer Person in die religiöse Gemeinschaft.
- b) Migration im Zusammenhang mit dem Wechsel (Übertritt) Angehöriger religiöser Gemeinschaften von einer Gemeinschaft in die andere.
- c) Migration als ökonomische, gemeinschaftspolitische oder persönliche Notwendigkeit im Kontext des spezifischen Lebenszusammenhangs in der religiösen Gemeinschaft<sup>47</sup>.

---

<sup>45</sup> Mitterauer: Probleme (s. Anm. 26).

<sup>46</sup> Als einen Überblick zur Verarbeitung mittelalterlicher Quellen und zu den dabei auftauchenden Problemen vgl. zuletzt den Sammelband A. Gilmour-Bryson (ed.): *Computer Applications to Medieval Studies* (=Studies in Medieval Culture 17), Kalamazoo, 1984. Zum Stand bis zum Beginn der achtziger Jahre vgl. C. Bourlet, Ch. Doutrelepon u. S. Lusignan: *Ordinateur et études médiévales. Bibliographie I*, Montreal, 1982.

<sup>47</sup> Da aufgrund vorhandener Quellen und Literatur die soziale Herkunft der Mitglieder einer Reihe von Konventen eruierbar ist, läßt sich innerhalb dieser Gruppen der Einfluß schichtrelevanter Merkmale auf das Migrationsverhalten relativ gut testen, obwohl diese nicht im Vordergrund stehen. Für manche Orden – im österreichischen Raum des Spätmittelalters vorrangig für Zisterzienser – konnten starke Übertrittsbewegungen von einem Konvent zu einem anderen festgestellt werden. Allein die bis jetzt publizierten Beispiele, die österreichische Zisterzienserklöster betreffen, belegen, daß jenes "Gyrovagantum" spätmittelalterlicher Klosterangehöriger große Teile Europas umspannt: vgl. G. Jaritz: *Cistercian Mi-*

## II. Studenten

### Quellen: Universitätsmatriken

Gemäß der räumlichen Beschränkung des Projekts werden vor allem Studenten der Stadt Wien behandelt. Wegen des guten Quellenkontinuums sind die Wiener Universitätsmatriken<sup>48</sup> für dieses Projekt eine besonders gut geeignete Quelle. Wenn die soziale Zusammensetzung der Studentenschaft auch nicht immer sehr gut bestimmbar ist, lassen sich durch das Herausfiltern bestimmter Gruppen (etwa aller Studenten aus Städten einer bestimmten Größe) "Aktivierungsindikatoren" für bestimmte Schichten gewinnen, die für längere Zeitspannen Gültigkeit haben. Relativiert werden kann die Untersuchung der Studenten an der Universität Wien durch die Heranziehung der österreichischen Studenten an nicht-österreichischen Universitäten<sup>49</sup>.

### III. Handwerker und Handwerksgesellen

Quellen: Rechnungsbücher, Urkunden, Bürgerbücher, Gesellen- und Meisterbücher

Die "klassischen" Migranten<sup>50</sup> des späten Mittelalters sind quellenmäßig in

---

grations in the Late Middle Ages, In: *Goad and Nail* (=Studies in Medieval Cistercian History 10 = Cistercian Studies Series 84), Kalamazoo, 1985, 191-200. Die systematische Untersuchung dieser Phänomene ist ein Desiderat ganz verschiedener Bereiche historischer Forschung. Im gleichen Maße über den engen regionalen Bereich hinausgehend erweisen sich Wanderungen von Angehörigen religiöser Institutionen im Zusammenhang mit dem klösterlichen Totengedenken, mit wirtschaftlichen Notwendigkeiten oder mit gemeinschaftspolitischen Angelegenheiten. Jene Art von "temporärer" Migration ist in der historischen Forschung bis jetzt noch weitgehend unberücksichtigt geblieben, wenn man von einigen Untersuchungen von Totenroteln absieht.

<sup>48</sup> F. Gall, W. Szaivert u.a. (Bearb.): *Die Matrikel der Universität Wien* (=Publikationen des Instituts für Österreichische Geschichtsforschung, VI. Reihe, 1. Abt.), Bd. 1 - 5, Wien / Köln / Graz, 1956 - 1975.

<sup>49</sup> Vgl. etwa A. Luschin v. Ebengreuth: *Österreicher an italienischen Universitäten zur Zeit der Rezeption des römischen Rechts*, in: *Blätter des Vereins für Landeskunde von Niederösterreich* 14 (1880), 228 - 252, 401 - 420, 15 (1881), 83 - 113, 250 - 264, 379 - 402, 417 - 428, 16 (1882), 64 - 72, 236 - 273, 17 (1883), 393 - 411, 490 - 516, 18 (1884), 271 - 316, 431 - 446,, 19 (1885), 503 - 558.

<sup>50</sup> Vgl. W. Reininghaus: *Die Migration der Handwerksgesellen in der Zeit der Entstehung ihrer Gilden*, in: *VSWG* 68 (1981), 1 - 21; - R.S. Elkar: *Umriss einer Geschichte der Gesellenwanderung im Übergang von der frühen Neuzeit zur Neuzeit*, in: *ders.* (Hrsg.): *Deutsches Handwerk in Spätmittelalter und früher*

dieser Zeit besonders schwer faßbar. Von wenigen listenartigen Quellen abgesehen, ist das Projekt auf die Sammlung (zahlreicher) Einzelbelege angewiesen. Die Differenzierung von Gesellenwanderung und "Experten"-wanderung entspricht verschiedenen Wandertypen. Während die zirkulären Wanderungen der Handwerksgelesen den seit etwa dem 14. Jahrhundert festgelegten Normen entspricht, liegt der Expertenwanderung, z.B. der Berufung eines Meisters in ein Kloster oder in eine Stadt, um genau abgegrenzte Arbeiten durchzuführen, eine Marktsituation zugrunde.

#### IV. Städtebürgertum

Quellen: Bürgerbücher, Rechnungsbücher, Urkunden

Die zeitlich differenzierten Zuzüge in spätmittelalterliche Städte gehören wahrscheinlich zu den besonders interessanten gesamtgesellschaftlichen Indikatoren für das Spätmittelalter und die frühe Neuzeit. Dies war einer der Hauptgründe für die Auswahl jener Gruppe. Die Hauptquelle für dieses Phänomen (Bürgerbücher) sind für den österreichischen Raum allerdings nur punktuell erhalten. Ein weiterer Aspekt ist hier die Entwicklung der temporären Migration vor allem im Zusammenhang mit interregionalem Handel.

#### V. Bergleute

Quellen: Urkunden, Rechnungsbücher

Diese Gruppe wurde vor allem als Kontrollgruppe für bestimmte Überlegungen sowie als typische Beispiel temporärer und nicht-temporärer Expertenmigration ausgewählt.

#### VI. Frauen

Quellen: Urkunden, Rechnungsbücher

"Frauen" unter soziale Gruppen zu subsumieren, mag (zunächst) verfehlt erscheinen. Die Geschichtswissenschaft beginnt erst in jüngster Zeit, sich mit dem Problem "Frauen" zu beschäftigen. Ob das Merkmal "weibliches Geschlecht" eine soziale Kategorie sei oder nicht, und welche Schlüsse dies gegebenenfalls zuließe, darüber ist die Diskussion, die von recht verschiedenen Gesichtspunkten aus geführt wird, noch im Gang<sup>51</sup>. In diesem Projekt wird davon ausgegangen, daß es beim heutigen Stand der wissenschaftlichen Diskussion notwendig ist, gesondert Material zum Problem "Mobilität und Frauen" zu erheben, um projektintern Vergleichsmöglichkeiten zu schaffen und darüber hinaus zur Gesamtdiskussion beitragen zu können.

---

Neuzeit (=Göttinger Beiträge zur Wirtschafts- und Sozialgeschichte 9), Göttingen, 1984, 85 - 116.

<sup>51</sup> Vgl. Ch. Klapisch-Zuber: *Le médiéviste, la femme et le sériel*, in: M. Perott (éd.): *Une histoire des femmes est-elle possible?*, Marseille, 1984, 37-47.

B. Auf einer zweiten Ebene wird der zu behandelnde Personenkreis im Hinblick auf seine Zugehörigkeit zu sozialen Schichten untersucht. So umstritten in der sozialgeschichtlichen Forschung das Konzept sozialer Schichtung im Zusammenhang mit mittelalterlichen und frühneuzeitlichen Gesellschaften sein mag, erscheint es doch als Arbeitsbasis brauchbar. Führen die zu analysierenden Quellengruppen ökonomische Daten an, so werden sie nach Möglichkeit zur Stratifizierung verwendet<sup>52</sup>.

In vielen Fällen liegen allerdings quantitativ verwertbare ökonomische Daten nicht vor. Dann wird auf die in der traditionellen Forschung weithin akzeptierte Methode der Klassifizierung auf der Basis von im gegebenen Zeitraum gebräuchlichen Anreden von Personen, Ämter- und Funktionsbezeichnungen, die in den Quellen auftreten, zurückgegriffen. Dies läßt gleichfalls eine Zuordnung im Sinn eines Schichtmodells zu. Dabei handelt es sich um ein rein hermeneutisches Verfahren, dessen Ergebnisse erst in einem weiteren Schritt quantitativen Methoden zugänglich sind.

\* \* \*

Ein zusätzlicher und gänzlich anderer Zugang zum Gesamtproblem wurde durch die Auswahl zweier "Ereignishorizonte" getroffen, die durch ihre Eigenart relativ gute Rückschlüsse auf mentalitätsgeschichtliche Prozesse zulassen: Unter Ereignishorizont im Sinne des Projekts, verstehen wir hier eine kohärente Dimension von Ereignissen, die Individuen unmittelbar betreffen, beziehungsweise die sie durch ihr unmittelbares Handeln bewirken. Diese Dimension ist dem Gesellschaftssystem inhärent, egal ob sich daraus systemstabilisierende oder -destabilisierende Tendenzen ergeben.

### I. Wallfahrten

Die Bedeutung der Wallfahrt für die Lebensgestaltung des mittelalterlichen Menschen wird in der historischen Forschung regelmäßig betont. Vom reichen überlieferten Quellenmaterial her abgeleitet wird die Wallfahrt des Mittelalters recht plakativ mit dem heute gängigen Begriff des Massentourismus in Verbindung gebracht. Systematische Untersuchungen zum konkreten Ausmaß jener Wanderbewegungen oder zum Einzugsbereich von Wallfahrtsorten fehlen für den österreichischen Raum vollständig<sup>53</sup>.

---

<sup>52</sup> Eine praktikable Methode zeigt E. Fügedi: Steuerliche Vermögen und soziale Gruppen in mittelalterlichen Städten, in: I. Batori (Hrsg.): Städtische Gesellschaft und Reformation. Kleine Schriften 2 (=Spätmittelalter und Frühe Neuzeit 12), Stuttgart, 1980, 58 – 96.

<sup>53</sup> Für England vgl. die inspirierende Arbeit von R.C. Finucane: Miracles and

## II. Kriminalität

Quellen: Urfehdebriefe, Richterrechnungen

Die Durchsicht größerer Bestände von Urfehdebriefen, die in der Vorbereitungsphase des Projekts durchgeführt wurde, hat gezeigt, daß deviantes Verhalten mit Migration häufig koinzidiert. Es scheint, daß durch den Ergebnishorizont der Kriminalität das Migrationsverhalten marginaler Schichten und Gruppen besonders gut zu erschließen ist.

### 5. Methoden und konkrete Vorgehensweisen

#### 5.1 Die Auswahl der Quellen

Bei der Quellenerhebung soll zunächst Vollständigkeit angestrebt werden. Die Einbeziehung einer Quelle in die Interpretation hängt davon ab, ob zunächst die konzeptuelle Grundeinheit, der "Einzelfall" mit seinen Merkmalen, hinreichend rekonstruierbar ist, dann davon, ob die Quelle im Kontext einer Subgruppe innerhalb des Konzepts der sozialen Gruppen und der Ereignishorizonte dem Projekt zuzuordnen ist. Ist dies nicht der Fall, wird die Quelle in einer "Residualgruppe" abgelegt, um sie zu einem späteren Zeitpunkt für eine gesonderte Interpretation heranzuziehen.

#### 5.2 Der Einzelfall als Analyseinheit

Als Grundeinheit der quantitativen Untersuchung wird aus den Quellen der Einzelfall einer Ortsveränderung (Wanderung) einer Person mit folgenden Informationen erhoben:

- Informationen zur Quelle und deren Kontext
- zum Anlaß, Typus und Kontext der Wanderung
- zur Person
- zum Herkunftsort
- zum Zielort.

#### 5.3 Zum Problem der Vergleichbarkeit: Das Konzept der Fallstudien

Die beobachteten Fälle von Wanderungen werden innerhalb des Konzepts der sozialen Gruppen und der Ereignishorizonte sinnvollen Sachgruppen, die klar definierten Raum-Zeit-Quellen-Entitäten entsprechen, zugeordnet. Die Zuwanderer nach Innsbruck, die den rechtlichen Status eines Inwohners erwarben, und die seit 1508 in einer listenartigen Quelle verzeichnet wurden, bilden zum Beispiel eine solche Subgruppe (vgl. unten Pkt. 7.2.1).

---

Pilgrims. Popular Beliefs in Medieval England, London / Melbourne / Toronto, 1977.

Über derartige Subgruppen können in einer ersten Phase der Interpretation im Rahmen von Fallstudien konkrete, gut überprüfbare Aussagen getroffen werden. Darüber hinaus können genaue Aussagen darüber gemacht werden, wie weit und wie gut eine Subgruppe mit einer anderen verglichen werden kann. Daraus ergibt sich nicht nur eine Möglichkeit zur Berücksichtigung historischer "Sonderphänomene", sondern vor allem ein hoher Grad an Reflektionsmöglichkeit für die eigenen Analysebedingungen. Die Gefahr voreiliger Verallgemeinerung wird stark reduziert.

## 5.4 Die formalisierte Behandlung der Quellen

In diesem Projekt kann keine Quellengattung grundsätzlich ausgeschlossen werden. Gegebene Quellenverluste müssen bei der Interpretation so berücksichtigt werden, daß eine Verzerrung der quantitativen Ergebnisse vermieden werden kann.

Innerhalb der Quellenformalisierung stellen die Formulierungen der Bedingungen zur Standardisierung der recht unterschiedlich strukturierten und unterschiedlichen Informationsgehalt aufweisenden Quellen das größte Problem dar<sup>54</sup>. Mit den Techniken der (traditionellen) empirischen Sozialwissenschaft, die weitgehend die Anlage "apriorischer" Codebooks und damit eine im gegenständlichen Fall unzulässige Einengung der Analysemöglichkeiten auf der Basis rein numerischer Datensätze verlangt, wäre diese Problematik mit Sicherheit nicht praktikabel lösbar.

Die Bearbeitung und Aufbereitung der Quellen mit dem Programmsystem CLIO<sup>55</sup> entschärften die Konflikte, die bei der apriorischen Anlage von Codebooks und einer darauffolgenden Erarbeitung des Quellenmaterials entstehen. Dieses datenbankorientierte System erlaubt es, die Quellsituation beliebig genau – im Migrationsgeschichteprojekt *sehr* genau – zu erfassen und abzuspeichern, sowie maschinell jeweils neue Codebooks nach der Erfassung des Materials zu erstellen, die sich an jeweils neuen Fragestellungen orientieren können. Daraus ergibt sich ein quasi "interaktiver" Forschungsprozeß, der eine Reihe von Problematiken, wie sie sich aus der Erfahrung quantitativer Projekte ergeben haben, umgeht. Die Verarbeitung 'unscharfer' Daten,

---

<sup>54</sup> Vgl. M. Overton: Computer analysis of an inconsistent data source: the case of probate inventories, in: Journal of Historical Geography 3 (1977), 317 – 326.

<sup>55</sup> M. Thaller: CLIO. Ein datenbankorientiertes Programmsystem für Historiker. Vorabdruck der Kapitel 1 bis 3, Göttingen, 1982, Vorabdruck der Abschnitte 4.1 und 4.2, Göttingen, 1983; – ders.: CLIO. Einführung und Systemüberblick, Göttingen, 1983.

die durch CLIO gewährleistet ist<sup>56</sup>, – ungeachtet dessen, daß das Konzept der "fuzzy sets" in den mathematischen Wissenschaften nicht unumstritten ist – erlaubt es, eine (immer einschränkende) Standardisierung nicht von Anfang an festzulegen<sup>57</sup>.

Aus diesen technischen Gegebenheiten eröffnet sich eine reale Chance, den traditionellen Konflikt zwischen hermeneutischen und quantifizierenden Ansätzen – natürlich nicht forschungslogisch – in praxi weitgehend zu entschärfen, wenn nicht aufzuheben<sup>58</sup>.

## 6. Zur Dateneingabe

Um das vorgehend Angeführte zu illustrieren, soll kurz das für das Projekt erarbeitete Dateneingabeschema angeführt und erläutert werden, welches den im Projekt verfolgten Intentionen gemäß flexibel gehalten wurde. Das Datendesign folgt dem von CLIO vorgeschlagenen, weitgehend hierarchisch organisierten Dokumentensystem. Eine der für das Projekt wesentlichen Vorüberlegungen (Konzept der Fallstudien) geht in das Datendesign ein. Auf der niedrigsten Ebene der Datenhierarchie wird prinzipiell folgendes Submodell eines Migrationsaktes (Einzelfall) eingegeben<sup>59</sup>:

---

<sup>56</sup> M. Thaller: Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte 'unscharfer' Systeme, in: H. Nagl-Docekal u. F. Wimmer: Neue Ansätze in der Geschichtswissenschaft (=conceptus studien 1), Wien, 1984, 77 – 100 und ders.: The Winds of Change. Problems of a Databank Oriented System Using the Concept of Fuzzy Sets, Paper Presented at the 1981 Joint Conference of IFDO and IASSIST, Grenoble.

<sup>57</sup> Ähnliche Probleme sind in der internationalen Forschung bei der Bearbeitung von Nachlaßinventaren aufgetaucht, wo eine Reihe von verschiedenen Lösungsmöglichkeiten erarbeitet wurde. Vgl. Probate Inventories, Papers Presented at the Leeuwenborch Conference (=A.A.G. Bijdragen 23), Wageningen, 1980.

<sup>58</sup> Vgl. dazu M. Thaller: Zur Formalisierbarkeit hermeneutischen Verstehens in der Historie, in: Mentalitäten und Lebensverhältnisse. Beispiele aus der Sozialgeschichte der Neuzeit. Rudolf Vierhaus zum 60. Geburtstag, Göttingen, 1982, 439 – 454; die von Thaller hier vertretenen Ansätze entsprechen den neueren Entwicklungen in den Sozialwissenschaften; vgl. M. Pollak: From Methodological Prescription to Socio-historical Description. The Changing Metascientific Discourse, in: Fundamenta Scientiae 4 (1983), 1 – 27.

<sup>59</sup> Die einzelnen inhaltlichen Kategorien werden durch '/' getrennt. A\$ steht für "Migrationsakt", P\$ für Person, HO\$, ZO\$ für Herkunfts- bzw. Zielort, NB für evtl. relevante Nebenbemerkungen. Unter "Typus" werden die Probleme tem-



A\$Typus/Quellengattung/Datum/Ursache,Motiv

P\$Geschlecht/Vorname/Nachname/Beruf/Stand/Anrede/Funktion/NB

HO\$Ort/Bezirk/Land/Bemerkungen über einen früheren

Herkunftsort

ZO\$Ort/Bezirk/Land

Dieser 'harte' vierzeilige Kern, den eigentlich aus den Quellen erhobenen individuellen Migrationsakt betreffend, ist zwingend notwendiger Bestandteil der den Quellen jeweils angepaßten Datenstruktur.

Für die Aufnahme von Urkunden ist über diesen Kern hinausgehend etwa folgendes Schema zur einheitlichen Datenaufnahme vorgesehen<sup>G1)</sup>:

D\$Datum des Dokuments/Bezeichnung/Archiv, Lit./Bemerkung

R\$Regest

DP\$Geschlecht/Vorn./Nachn./Beruf/Stand/Anrede/Funktion

A\$Typus/Quellengattung/Datum/Ursache,Motiv

P\$Geschlecht/Vorname/Nachname/Beruf/Stand/Anrede/  
Funktion/NB

HO\$Ort/Bezirk/Land/Bem. über einen früheren

Herkunftsort

ZO\$Ort/Bezirk/Land

Für listenartige Dokumente (vor allem Rechnungsbücher) gelten folgende Regeln:

Wenn in einer listenartigen Quelle mehr als 20 % der Einzeleintragungen sich auf Migrationsakte beziehen, ist die gesamte Einzelquelle maschinenlesbar zu machen. In einem individuellen hierarchischen Datenmodell wird die

---

porärer und nichttemporärer Migration erfaßt. Unter "Ursache, Motiv" werden die Informationen, die die Einzelquelle bietet, aufgenommen (Illustratives Beispiel: Empfehlungsschreiben eines Abtes für die Aufnahme eines Mönches in einem anderen Kloster nach dem Brand des Stammklosters). Die Einrückungen zeigen die Hierarchisierung im Datenmodell an. Modellbestimmende Attribute von HO und ZO, etwa die zentralörtliche Funktion vs. nichtzentralörtliche Funktion, Größe der Orte, Entfernung von HO nach ZO werden auf der Ebene der maschinellen Datenerfassung erst in der Evaluationsphase eingefügt.

<sup>G1)</sup> D\$ steht für Dokumentation (der Quelle, in diesem Falle der Urkunde), R\$ für Regest bzw. nähere Beschreibung des Inhalts, DP\$ für die ausstellende Person der Quelle.

Struktur der einzelnen Quelle abgebildet. Das Grundmodell dafür sieht so aus<sup>G1</sup>:

```
D$Datum des Dokuments/Bezeichnung/Archiv, Lit./Bemerkung
R$Beschreibung der Struktur und der Besonderheiten
  EIN$Einnahmen
    I$....
    I$....
    A$Typus/Quellengattung/Datum/Ursache,Motiv
    P$Geschlecht/Vorname/Nachname/Beruf/Stand/Anrede/
      Funktion/NB
      HO$Ort/Bezirk/Land/Bem. über einen früheren
      Herkunftsort
      ZO$Ort/Bezirk/Land
    I$....
    I$....
  AUS$Ausgaben
    I$....
    I$....
    A$Typus/Quellengattung/Datum/Ursache,Motiv
    P$Geschlecht/Vorname/Nachname/Beruf/Stand/Anrede/
      Funktion/NB
      HO$Ort/Bezirk/Land/Bem. über einen früheren
      Herkunftsort
      ZO$Ort/Bezirk/Land
    I$....
```

Die meisten dieser Quellen sind nach Ein- und Ausgaben getrennt angelegt. Die einzelnen Eintragungen enthalten mehrere, je nach dem Gegenstand, den die Quelle betrifft, differierende Detailinformationen, die unter I\$ systematisch eingetragen werden sollen. Enthält eine Einzeleintragung Informationen über einen Migrationsakt, wird dieser auf einer hierarchisch niedrigeren Ebene mit Hilfe des oben beschriebenen Submodells festgehalten (vgl. Beispiel 7.2.4).

Auf diese Weise erscheint das Konzept der Fallstudien am leichtesten realisierbar. Jede Einzelquelle kann für sich untersucht werden, die Migrationsakte können bei der vorgesehenen Datenorganisation "leicht" aus beliebigen Quellen zusammengestellt und miteinander verglichen werden. Ferner

---

<sup>G1</sup> EIN\$ und AUS\$ stehen für die nähere Beschreibung der Einnahmen bzw. Ausgaben, I\$ für Informationen zu den Inhalten der Einzeleintragungen.

bietet CLIO einfache Möglichkeiten zur Anlage verschiedener Formen von Thesauri und zur Umwandlung des CLIO-Datensatzes in SPSS-Datensätze, welche die breite Palette der Anwendung statistischer Methoden eröffnen. Darüber hinaus bestehen natürlich relativ leicht Möglichkeiten, die für das Projekt angelegten Datensätze für Sekundäranalysen zu anderen Fragestellungen zu verwenden<sup>62</sup>.

## 7. Beispiele

Anhand einiger Beispiele wollen wir zeigen, wie dieses Modell in die praktische Arbeit mit unterschiedlichen Quellen umgesetzt wird. Die einzelnen Beispiele sind jeweils kurze Ausschnitte aus größeren Datenbeständen.

### 7.1. Urkunden

**7.1.1 Empfehlungsschreiben von Äbten zisterziensischer Konvente an die Vorsteher anderer Konvente für zu diesen gesandte Mönche<sup>63</sup>:**

D\$1445-1453/Empfehlungsschreiben/UB-Graz, Hs. 962, fol. 151v-152r  
 R\$Empfehlungsschreiben des Abtes von Lilienfeld an den Abt von  
 Neuberg fuer einen Professoren von Sedlec  
 DP\$M/Petrus//Moench/Regularklerus//Abt von Lilienfeld  
 A\$Klosterwanderung/Empfehlungsschreiben/1445-1453  
 P\$M/Petrus//Moench%Profess von Sedlec/Regularklerus  
 HO\$Lilienfeld#Zisterzienserkloster/Lilienfeld/Niederoesterreich  
 ZO\$Neuberg#Zisterzienserkloster/Nuerzzuschlag/Steiermark

<sup>62</sup> Zur Sekundäranalyse maschinenlesbarer literarischer Texte des Mittelalters vgl. M. Thaller: Recycling the Drudgery. On the Integration of Software Supporting Secondary Analysis of Machine-Readable Texts into a DBMS, In: L. Cignoni and C. Peters (ed.): Computers in Literary and Linguistic Research (=Linguistica Computazionale 3 Supplement), Pisa, 1983, 253-268; – G. Jaritz, Daily Life in Medieval Literature, in: MAQ-Newsletter 2 (1984), 6-23.

<sup>63</sup> Zur besseren Verständlichkeit sei das ausführliche Regest des ersten Falles angegeben. (1445-1453), Lilienfeld: Abt Petrus von Lilienfeld übersendet dem Abt Johannes von Neuberg den Bruder Petrus, Konventualen des Klosters Sedlec, der von dessen Abt Paul zu ihm geschickt worden war, und bittet, diesen in die Neuberger Klostersgemeinschaft aufzunehmen. – Die im Datensatz vorkommenden '%' bezeichnen eine darauffolgende verbale Kategorie zur näheren Erläuterung, '#' zeigt an, daß nach der Ortsangabe eine nähere Eingrenzung einer Institution, etc. erfolgt.

A\$Klosterwanderung//vor 1445-1453/Hussiten ?  
P\$M/Petrus//Moench%Profess von Sedlec/Regularklerus  
HO\$Sedlec#Zisterzienserkloster/Pardubitz/Boehmen  
ZO\$Lilienfeld#Zisterzienserkloster/Lilienfeld/Niederoesterreich  
D\$1454-1460/Empfehlungsschreiben/UB-Graz, Hs.962, fol.164r-165v  
R\$Empfehlungsschreiben des Abtes von Neukloster an den Abt von  
Neuberg fuer einen Professoren von Rein  
DP\$M/Gottfried//Moench/Regularklerus//Abt von Neukloster  
A\$Klosterwanderung/Empfehlungsschreiben/1454-1460/  
Ueberfuellung des Klosters  
P\$M/Heinrich//Konverse%Profess von Rein/Konverse  
HO\$Wiener Neustadt#Zisterzienserkloster Neukloster/Wiener  
Neustadt/Niederoesterreich  
ZO\$Neuberg#Zisterzienserkloster/Muerzzuschlag/Steiermark  
A\$Klosterwanderung//vor 1454-1460  
P\$M/Heinrich//Konverse%Profess von Rein/Konverse  
HO\$Ebrach#Zisterzienserkloster/Bamberg/Franken  
ZO\$Wiener Neustadt#Zisterzienserkloster Neukloster/Wiener  
Neustadt/Niederoesterreich  
A\$Klosterwanderung//vor 1454-1460  
P\$M/Heinrich//Konverse%Profess von Rein/Konverse  
HO\$Rein#Zisterzienserkloster/Graz-Umgebung/Steiermark  
ZO\$Ebrach#Zisterzienserkloster/Bamberg/Franken

### 7.1.2 Urfehdebriefe

D\$28.06.1399/Urfehdebrief/StA Krems, Urk. 100  
R\$Michael Tell von Aggsbach und sein Frau Margret, die wegen  
eines Diebstahls in Langenlois verhaftet und in das Gericht  
nach Krems gebracht wurden, schworenen, nachdem sie von der  
Todesstrafe begnadigt worden sind, Urfehde.  
DP\$M/Michael/Tell  
DP\$F/Margret/Tell%Frau des Michael T.  
A\$ungeklaert%Verhaftung wegen eines Diebstahls/Urfehdebrief/vor  
28.06.1399/ungeklaert  
P\$M/Michael/Tell  
HO\$Aggsbach/Krems%oder Melk - beide Varianten moeglich!/  
Niederoesterreich  
ZO\$Langenlois/Langenlois/Niederoesterreich  
P\$F/Margret/Tell

HO\$Aggsbach/Krems%oder Melk - beide Varianten moeglich!/  
Niederoesterreich  
ZO\$Langenlois/Langenlois/Niederoesterreich

## 7.2. Listenförmige Quellen

### 7.2.1 Inwohneraufnahmen in die Stadt Innsbruck<sup>64</sup>

D\$1508-1567/Inwohneraufnahmeverzeichnis/K. Schadelbauer u.  
M. Fritz, Die Innsbrucker Inwohneraufnahmen von 1508 bis 1567,  
Veroeff. aus dem Stadtarchiv Innsbruck 26, Innsbruck, 1964, 5  
R\$seit 1508 wurde in Innsbruck neben dem Buergerbuch ein  
Inwohneraufnahmeverzeichnis gefuehrt, das die Inwohneraufnahme  
Person fuer Person verzeichnet und (fast) jede Person unter  
Nennung des zu zahlenden Geldbetrages nennt und wahrscheinlich  
fuer die Abrechnung des Stadtschreibers gedient hat.  
A\$Inwohneraufnahme/Inwohneraufnahmeverzeichnis/20.03.1509/  
Inwohneraufnahme  
P\$M/Jorg/Enndorfer//Inwohner  
HO\$Enndorf ?  
ZO\$Innsbruck/Innsbruck/Tirol  
A\$Inwohneraufnahme/Inwohneraufnahmeverzeichnis/31.03.1509/  
Inwohneraufnahme  
P\$M/Ulrich/Spaet/Schmied/Inwohner  
HO\$Kiebach%OG. Hofkirchen im Traunkreis ?/Linz-Land/  
Oberoesterreich  
ZO\$Innsbruck/Innsbruck/Tirol  
A\$Inwohneraufnahme/Inwohneraufnahmeverzeichnis/11.05.1509/  
Inwohneraufnahme  
P\$M/Hans/Wolman//Inwohner  
HO\$Landshut/Landshut/Bayern  
ZO\$/Innsbruck/Innsbruck/Tirol

<sup>64</sup> Zur Erläuterung sei der letzte, im Beispiel angegebene Fall in jener Form wiedergegeben, in welcher er in der Edition auftritt: *Schadelbauer - Fritz: Inwohneraufnahme* 5, n. 10: "1509 Mai 11. Hans Wolman von Landhuet zu iw. aufg."

### 7.2.2 Universitätsmatrikel<sup>65</sup>

D\$1451-1518/Universitätsmatrikel/F. Gall, Die Matrikel der  
Universitaet Wien (Publikationen des Instituts fuer  
Oesterreichische Geschichtsforschung, Reihe VI, Abt. 1, Bd. 2)  
Wien, 1959, 94

A\$Immatrikulation/Universitätsmatrikel/13.10.1466/Studium

P\$M/Sigismundus/Collrer//Student///Gr=4

HO\$Radstadt/Radstadt/Salzburg

ZO\$Wien#Universitaet/Wien/Wien

A\$Immatrikulation/Universitätsmatrikel/13.10.1466/Studium

P\$M/Johannes/Firstain//Student///Gr=4

HO\$Radkersburg/Radkersburg/Steiermark

ZO\$Wien#Universitaet/Wien/Wien

A\$Immatrikulation/Universitätsmatrikel/13.10.1466/Studium

P\$M/Andreas/Purckhollczler//Student///Gr=4

HO\$Eferding/Eferding/Oberoestereich

ZO\$Wien#Universitaet/Wien/Wien

A\$Immatrikulation/Universitätsmatrikel/13.10.1466/Studium

P\$M/Andreas/Pulsinger//Student///Gr=4

HO\$Senftenberg/Krems/Niederoesterreich

ZO\$Wien#Universitaet/Wien/Wien

A\$Immatrikulation/Universitätsmatrikel/13.10.1466/Studium

P\$M/Martinus/Pileatoris//Student///Gr=2

HO\$Wien/Wien/Wien

ZO\$Wien#Universitaet/Wien/Wien

### 7.2.3 Meister- und Gesellenverzeichnis<sup>66</sup>

D\$1489/MEISTER- UND GESELLENBUCH DER KREMSER ZIMMERLEUTE/

STA-KREMS, ZUNFTARCHIV 29,392, FOL.26R

A\$HANDWERK/MEISTERBUCH/1489/MITGLIEDSCHAFT IN DER BRUDERSCHAFT  
DER ZIMMERLEUTE

P\$M/MICHEL/SECKAWER/ZIMMERMANN///MEISTER

HO\$HERZOGENBURG/HERZOGENBURG/NIEDEROESTERREICH

<sup>65</sup> Letzter Fall im Beispiel: Gall - Szaivert: Matrikel 94: "Martinus Pileatoris de  
Wienna 2 gr."

<sup>66</sup> Erster Fall im Beispiel: Meisterbuch fol.26v: "Vermerkcht was maister bey  
unns in der bruederschaft sein herin aus dem land, die dann auch mit namen hernach  
volgent: ... Maister Michel Seckawer von Herczogburgk; ....".

ZO\$KREMS/KREMS/NIEDEROESTERREICH  
 A\$HANDWERK/MEISTERBUCH/1489/MITGLIEDSCHAFT IN DER BRUDERSCHAFT  
 DER ZIMMERLEUTE  
 P\$M/MICHEL/KEFFERNÜLLNER/ZIMMERMANN///MEISTER  
 HO\$ST.POELTEN/ST.POELTEN/NIEDEROESTERREICH  
 ZO\$KREMS/KREMS/NIEDEROESTERREICH  
 A\$HANDWERK/MEISTERBUCH/1489/MITGLIEDSCHAFT IN DER BRUDERSCHAFT  
 DER ZIMMERLEUTE  
 P\$M/HANS/GERHART/ZIMMERMANN///MEISTER  
 HO\$WILHELMSBURG/ST.POELTEN/NIEDEROESTERREICH

## 7.2.4 Rechnungen<sup>67</sup>

D\$1495-1496/RECHNUNGSBUCH DES GOETTWEIGER ABTES/STIA.GOETTWEIG  
 R\$DAS RECHNUNGSBUCH ENTHAELT IN DETAILLIERTER AUFSCHLUESSELUNG  
 UND STRUKTURIERUNG DIE EINNAHMEN UND AUSGABEN DES ABTES VON  
 GOETTWEIG  
 AUS\$AUSGABEN FUER DIE REISEN DES ABTES  
 I\$VON WEGEN HANF 7 S. 23 D.  
 A\$EINKAUFSREISE/RECHNUNGSBUCH/9.1.1495/HANFKAUF  
 P\$M/MATTHIAS//MOENCH/REGULARKLERUS//ABT VON GOETTWEIG  
 HO\$GOETTWEIG#BENEDIKTINERSTIFT/KREMS/NIEDEROESTERREICH  
 ZO\$KREMS/KREMS/NIEDEROESTERREICH  
 I\$VON WEGEN STEUER UND ANDERER NOTDURFT 14 TAGE 7 TL. 11 D.  
 A\$DIENSTREISE/RECHNUNGSBUCH/2.2.1495/STEUERANGELEGENHEITEN  
 P\$M/MATTHIAS//MOENCH/REGULARKLERUS//ABT VON GOETTWEIG  
 HO\$GOETTWEIG#BENEDIKTINERSTIFT/KREMS/NIEDEROESTERREICH  
 ZO\$WIEN/WIEN/WIEN  
 I\$VON WEGEN DES RUEBER SACH 60 D.  
 A\$DIENSTREISE/RECHNUNGSBUCH/8.3.1495/RECHTSSACHE ?  
 P\$M/MATTHIAS//MOENCH/REGULARKLERUS//ABT VON GOETTWEIG  
 HO\$GOETTWEIG#BENEDIKTINERSTIFT/KREMS/NIEDEROESTERREICH  
 ZO\$STEIN/KREMS/NIEDEROESTERREICH

<sup>67</sup> Auszug aus der Rechnung: Rechnungsbuch des Abtes 1495/95, fol.18r: "Auff des herren rais: Item am freytag nach sand Erhardt tag gen Krembs von wegen des haniff 7 s. 23 d., ... Item secunda feria post Plasy gen Wien von wegen der steur und ander notturfft verzertt 14 tag 7 tl. 11 d., ....."

### 7.3 Mirakelverzeichnis<sup>68</sup>

D\$1518-1522/VOTIVBILD/GRAZ, STM. LANDESMUS. JOANNEUM  
R\$VOTIVBILD DES VOM BLITZ GETROFFENEN HANS PODSACHK AUS MAEHREN,  
DER DURCH ANRUFUNG DER GOTTESMUTTER VON MARIAZELL ERRETTET  
WIRD.

A\$WALLFAHRT ?%KEINE BEMERKUNG ZUR TATSAECHLICHEN AUSFUEHRUNG/  
VOTIVBILD/VOR 1518-1522/BLITZSCHLAG

P\$M/HANS/PODSACHK

HO\$//MAEHREN

ZO\$MARIAZELL/MARIAZELL/STEIERSMARK

R\$VOTIVBILD EINER FRAU AUS BOEHEINKIRCHEN, DIE IM KINDBETT  
GELAEHMT UND DURCH ANRUFUNG DER GOTTESMUTTER VON MARIAZELL  
ERRETTET WIRD.

A\$WALLFAHRT %KEINE BEMERKUNG ZUR TATSAECHLICHEN AUSFUEHRUNG/  
VOTIVBILD/VOT 1518-1522/LAEHMUNG IM KINDBETT

P\$F

HO\$BOEHEINKIRCHEN/ST. POELTEN/NIEDEROESTERREICH

ZO\$MARIAZELL/MARIAZELL/STEIERSMARK

### 7.4 Belege aus der Sekundärliteratur (Klostereintritte in das Zisterzienserkloster Rein)<sup>69</sup>

D\$26.7.1427/URKUNDE/RA-URK. A IX 16; G. JARITZ, DIE  
KONVENTUALEN

<sup>68</sup> Erster Fall im Beispiel, Bildlegende: "Hans Podsachk ausz Merhern von einem donnerschlag und himlicz ward erschlagen und lang dot gelegen. So bald er gen Zell zu unser frauen versprechen, ward er gesunt". - Hinsichtlich der Migrationsproblematik handelt es sich hierbei um Grenzfälle, da die Anrufung der Muttergottes von Mariazell natürlich nicht unbedingt mit einer diesbezüglichen Wallfahrt verbunden sein muß. Nichtsdestoweniger erscheint die Aufnahme solcher Belege wichtig, um Ergänzungen bzw. Relativierungen zu tatsächlich belegten Wallfahrten zu erhalten.

<sup>69</sup> Bei manchen Umsetzungen, vor allem solchen, die aus der Sekundärliteratur stammen, fehlen vollständige Informationen, wie sie Originalquellen liefern. Nachdem es aber bei guten Aufbereitungen sinnlos erscheint, gesamte Urkundenbestände nochmals heranzuziehen, wird auf Teile, die nicht direkt den Migrationsakt betreffen, bewußt verzichtet, in diesem Falle etwa auf DP\$ der Urkunden. - "" bedeutet, daß der Beleg keinen Bezug auf den Migrationsakt an sich hat, daß jedoch durch in ihm vorhandene Angaben (in diesem Fall der Nachname) auf Herkunft und Migrationsakt (in diesem Fall Klostereintritt) geschlossen werden kann.



DER ZISTERZEN REIN, SITTICH UND NEUBERG IM MITTELALTER. PHIL. D  
ISS. GRAZ, 1973, II. BD., 20 F.

A\$KLOSTEREINTRITT\*/URKUNDE/26.7.1427

P\$M/JOHANNES/SAILER/MOENCH/REGULARKLERUS:BUERGER\*//ABT 1425-  
1428

HO\$GRAZ/GRAZ/STEIERSMARK

ZO\$REIN#ZISTERZIENSERKLOSTER/GRAZ-UMGEBUNG/STEIERSMARK

D\$17.12.1438/URKUNDE/RA-URK. A IX 23; G. JARITZ, DIE  
KONVENTUALEN DER ZISTERZEN REIN, SITTICH UND NEUBERG IN . PHIL.  
MITTELALTER PHIL. DISS. GRAZ, 1973, II. BD., 23

A\$KLOSTEREINTRITT\*/URKUNDE/17.12.1438

P\$M/JOHANNES/AICHSTAETTER/MOENCH/REGULARKLERUS//ABT 1433-1439

HO\$EICHSTAETT/EICHSTAETT/BAYERN

ZO\$REIN#ZISTERZIENSERKLOSTER/GRAZ-UMGEBUNG/STEIERSMARK

## 7.6 Sekundäranalyse maschinenlesbarer Quellen

Als Beispiel für die Möglichkeit einer Sekundäranalyse maschinenlesbarer Quellenaufbereitung sei das Beispiel der Datenbank des Instituts für mittelalterliche Realienkunde Österreichs der Österreichischen Akademie der Wissenschaften angeführt. Unter Anwendung von CLIO werden dort Beschreibungen mittelalterlicher Bildquellen unter besonderer Berücksichtigung ihres Aussagegehaltes zu Alltag und materieller Kultur des Mittelalters computergestützt durchgeführt. Ohne auf das Beschreibungsschema und auf die praktischen und technischen Möglichkeiten der Durchführung der Sekundäranalyse eingehen zu können, sei hier ein Beispiel gegeben, das einen Migrationsakt enthält (die betreffenden Quellenteile sind durch Unterstreichung hervorgehoben) und damit für unsere Fragestellungen relevant wird<sup>64</sup>.

---

<sup>64</sup> Zu den angeführten Bildbeschreibungen und dem dabei angewandten System, das inzwischen in gewissen Einzelheiten Modifizierungen erfahren hat, vgl. G. Jaritz: *Daily Life in the Middle Ages, Iconography of Medieval Art and the Use of EDP*, in: *Historical Social Research* 21 (1981); - ders.: *EDV-Anwendung in der Ikonographie des Mittelalters*, in: Bericht über den fünfzehnten österreichischen Historikertag in Salzburg (=Veröffentlichungen des Verbandes österreichischer Geschichtsvereine 23), Wien, 1984, 373-384; - E. Vavra: *Möglichkeiten einer EDV-unterstützten Auswertung mittelalterlicher Bildquellen*, in: *Beiträge zur Überlieferung und Beschreibung deutscher Texte des Mittelalters* (Göppinger Arbeiten zur Germanistik 402) Göppingen, 1983, 195-211. Zur Vorgängerversion vgl. ausführlicher M. Thaller: *DESCRIPTOR* (s. Anm. 4), 166-194.

N\$00165-%OHNE GESAMTAUFNAHME/7002505-7002506/3005125;3005127  
 G\$TAFELMALEREI/PREDELLA EINES FLUEGELALTARS//STIFTERFAMILIE/  
 NIEDEROESTERREICH  
 D\$1515/1528/NEUKIRCHEN AM OSTRONG/NIEDEROESTERREICH/PFARRKIRCHE  
 C\$/NEUKIRCHEN AM OSTRONG%NIEDEROESTERREICH  
 B\$BETEN  
 S\$M/SCHAUHINGER/STIFTER;NIEDERADELIGER%BEI ST. POELTEN,  
 NIEDEROESTERREICH/KNIEND  
 T\$HAARTRACHT/GRAU//SCHULTERLANG;GLATT  
 A\$BIRETT/BLAUGRAU/SEIDE/GLAENZEND  
 T\$KREMPE/BRAUN/PELZ/AUFGEBOGEN  
 K\$SCHAUBE/WEINROT  
 T\$KRAGEN/BRAUN/PELZ  
 T\$UNBEKANNTER KLEIDERTEIL %SKAPULIERARTIG  
 S\$M/SCHAUHINGER/STUDENT%UNIVERSITAET WIEN;NIEDERADELIGER%BEI ST.  
 POELTEN;STIFTERSOHN%AELTETER/KNIEND  
 T\$HAARTRACHT/BLOND//SCHULTERLANG;GEWELLT  
 A\$ARMILLARSPHAERE/GOLD  
 K\$BIRETT/ROT//FLACH  
 K\$STUDENTENKLEID!/GRAU  
 K\$ROCK/GRAU  
 K\$SCHULTERMAENTELCHEN/GRAU  
 T\$SCHNUEERUNG/GELB  
 X\$ANHAENGER%ZWEI/ROT;GRUEN/STOFF(?)/TROPFENFOERMIG  
 T\$BORTE/GELB  
 K\$MANTEL%EINE HAELFTE GESCHULTERT/GRAU  
 T\$KLEIDERFUTTER/GELB



**Datenbeschreibung und Datendokumentation  
in der historischen Sozialforschung.**

**Problembeschreibung und Empfehlungen für die Forschung.**

**1. Vorbemerkung: Datenbeschreibung und Datendokumentation in der Praxis der sozialwissenschaftlichen Datenarchive**

Zentraler Bestandteil der sozialwissenschaftlichen Infrastruktur vor allem in Westeuropa, in den skandinavischen Ländern und in Nordamerika ist seit den sechziger und den siebziger Jahren ein Netz von Datenarchiven. Diese Einrichtungen archivieren für die sozialwissenschaftliche Forschung maschinenlesbare Daten, in erster Linie Forschungsdaten, und stellen sie für weitere Untersuchungen zur Verfügung. Anfänglich hauptsächlich mit der Archivierung und Bereitstellung von Umfragedaten befaßt, erfassen die sozialwissenschaftlichen Datenarchive inzwischen auch andere für die sozialwissenschaftliche Forschung relevante Datentypen, z.B. Daten amtlicher statistischer Zählungen. Historische Daten für eine sich sozialwissenschaftlich verstehende historische Forschung, für die historische Sozialforschung, gehören bislang nur zu einem geringen Teil zum Aufgabenbereich der sozialwissenschaftlichen Datenarchive. Aufgrund der Entstehungsbedingungen und auch der aktuellen Situation der meisten sozialwissenschaftlichen Datenarchive – Umfragedaten machen immer noch die Masse der archivierten und für Sekundäranalysen bereitgestellten Daten aus – ist die von anderen Datentypen verschiedene Problematik einer Archivierung und Bereitstellung sozialwissenschaftlich relevanter historischer Daten gar nicht oder nur kaum reflektiert worden. Das Zentrum für historische Sozialforschung hat im Rahmen seiner – von der Deutschen Forschungsgemeinschaft geförderten – Archivierung und Bereitstellung maschinenlesbarer historisch-sozialwissenschaftlicher Forschungsdaten Ansätze für ein Instrumentarium entwickelt, das die Erfordernisse an eine Archivierung und Bereitstellung sozialwissenschaftlich relevanter historischer Daten abdeckt.

Im vorliegenden Beitrag wird ein Teil dieser Bemühungen um die Schaffung eines datenspezifischen Instrumentariums aufgegriffen: die Empfehlungen für Datenbeschreibungs- und Datendokumentationselemente in der historischen Sozialforschung. Die dabei formulierten und bereits praktizierten Empfehlungen decken einen Datenbeschreibungs- und Datendokumentationsbedarf ab, der in Teilen Beschreibungs- und Dokumentationsanfordernissen der auf der Grundlage von Umfrage- und anderen Daten arbeitenden Sozialwissenschaften und der traditionellen Geschichtswissenschaft entspricht. Diese Verknüpfung von empirischen Sozialwissenschaften und Geschichts-

wissenschaft manifestiert sich darin, daß sich in der historischen Sozialforschung theoriegeleitete Forschungsfragen und die Verwendung von historischen Quellen, bzw. von prozeß-produzierten Daten miteinander verknüpfen. Die Beschreibungs- und Dokumentationserfordernisse der historischen Sozialforschung gehen insofern über den entsprechenden Bedarf der Geschichtswissenschaft hinaus, als in der historischen Sozialforschung Informationen aus historischen Quellen in theoretisch relevante Sachverhalte transformiert werden. Dementsprechend besteht nicht nur die Notwendigkeit, vergleichbar den in der Geschichtswissenschaft entwickelten Standards, Quellen aufzulisten und ein Quellenverzeichnis zu erstellen, sondern es ist ein Beschreibungs- und Dokumentationsbedarf abzudecken, der auf ein umfassenderes "Verständnis" der Daten abzielt.

Die sozialwissenschaftlichen Datenarchive verfahren bei der Datenbeschreibung und -dokumentation gemäß den Standards, die in der empirischen Sozialforschung üblich geworden sind, ohne daß allerdings in jedem Fall nach ihnen gehandelt wird. Diese Standards betreffen:

1. Die Beschreibung der Grundgesamtheit oder Population. Die Grundgesamtheit ist in der empirischen Sozialforschung die wichtigste Bezugsgröße, die Einheit, auf die hin Aussagen bzw. Verallgemeinerungen beabsichtigt sind. "Die Grundgesamtheit (häufig auch als Universum oder Population bezeichnet) kann sich etwa auf die Bevölkerung eines Staates oder einer Gemeinde beziehen, kann aber auch nach einer kategorialen Zugehörigkeit von Personen bestimmt werden: etwa Richter, Sportler einer bestimmten Disziplin oder Minister in Deutschland. Die Grundgesamtheit kann sich aber auch Kollektive oder Artefakte beziehen: z.B. auf Organisationen, wie Unternehmen der Stahlindustrie oder die Finanzverwaltung des Bundes und der Bundesländer. Artefakte können beispielsweise Bilder des 19. Jahrhunderts sein, die Zeitungsberichterstattung über eine gewisse Klasse von Ereignissen (z.B. die Dreyfusaffäre oder Berichterstattung über Wissenschaft), aber auch Dokumente, wie z.B. Personalakten von Behörden." (Heine von Alemann, *Der Forschungsprozeß*, Stuttgart: Teubner 1977, S. 74)
2. Die Beschreibung bzw. Dokumentation derjenigen Merkmale, über die im Verlauf der Datenerhebung quantitative Informationen erhoben werden. Idealerweise ist der Ausgangspunkt dieses Vorgangs die Umsetzung einer theoretisch abgeleiteten Forschungsfrage in meßbare Größen, die wiederum in die Formulierung eines adäquaten Erhebungsinstruments (z.B. Fragebogens) mit (abhängig von der "Aufmachung" des Erhebungsinstruments) komplementärem Codeplan überleitet.

Der Datenbeschreibungs- und Dokumentationsbedarf wird in der empirischen Sozialforschung in der Weise abgedeckt, daß einerseits eine Definition bzw. eine Beschreibung der Grundgesamtheit gegeben wird; daneben die Merkmalsdimensionen der jeweiligen Untersuchung durch eine Wiedergabe des Erhebungsinstruments (Fragebogen und Codeplan) dokumentiert werden. Die sozialwissenschaftlichen Datenarchive decken diesen Datenbeschreibungs- und Datendokumentationsbedarf ab, indem sie neben einer Wiedergabe des Original-Fragebogens und des Codeplans detaillierte Informationen über die Grundgesamtheit der betreffenden Studie und über die jeweils gezogene Stichprobe zur Verfügung stellen. Die Beschreibung von Grundgesamtheit und Stichprobe wird im Rahmen der sog. "Studienbeschreibung" vorgenommen, einem Instrument, das von verschiedenen westeuropäischen Datenarchiven gemeinsam zur Beschreibung und Dokumentation von Umfragedaten entwickelt worden ist. Mit der Wiedergabe des Original-Fragebogens und des Codeplans und gegebenenfalls einer Auflistung der Variablen werden in der Regel die Anforderungen an eine Beschreibung und Dokumentation der Merkmalsdimensionen von Umfragedaten als erfüllt angesehen. Dies nicht zuletzt deshalb, weil in der Sozialforschung auf der Grundlage von Umfragedaten die Umsetzung der theoretisch abgeleiteten Forschungsfrage in Meßoperationen, d.h. die Operationalisierung, idealiter identisch ist mit der Fragebogenkonstruktion. Zudem wird ein zusätzlicher Beschreibungs- und Dokumentationsbedarf auch deshalb nicht gesehen, weil davon ausgegangen wird, daß die Inhalte eines Fragebogens bzw. die Variablen aus dem zeitgenössischen Alltagsverständnis deutlich werden. Dies kann für die historischen Kontexte der Inhalte der historischen Sozialforschung nicht geltend gemacht werden, sind doch die Alltagskonnotationen der erforschten Zeit von denen der Zeit des Forschers verschieden. Diese Probleme stellen sich jedoch auch zunehmend für die Sozialforschung auf der Basis von Umfragedaten. In der empirischen Sozialforschung enthalten Codebücher über die Bezeichnung der Variablen und über die Bezeichnung bzw. das Etikett der Variablenausprägung normalerweise keine weiteren Informationen zum inhaltlichen "Verständnis" der Variablen. Die Wiedergabe der Variablenbezeichnung, in der Umfrageforschung mit der einzelnen Frage des Fragebogens identisch, wird häufig noch als ausreichend angesehen; entsprechen sich doch Fragebogenformulierung und Bezugssystem des Alltags. Dieses Selbstverständnis der Umfrageforschung wird schon brüchig, sobald Sekundäranalysen von Umfragen z.B. der frühen fünfziger Jahre durchgeführt werden sollen. Die Bedeutung einzelner Variablen aus Untersuchungen der damaligen Zeit ist nicht ohne weiteres transportabel, entspricht doch das Bezugssystem der damaligen

Zeit nicht mehr unserem. Die Alltagskonnotationen einer Fragebogenformulierung aus den fünfziger Jahren sind uns nicht mehr bekannt.

## **2. Zentrale Datenbeschreibungs- und Datendokumentationsprobleme in der historischen Sozialforschung**

### **2.1 Grundgesamtheit, Auswahlgesamtheit, Erhebungseinheit und Quelle in der historischen Sozialforschung**

In der historischen Sozialforschung stellt sich verstärkt das Problem, inwieweit Grundgesamtheit bzw. Zielgesamtheiten durch ein vorhandenes Quellenmaterial adäquat abgebildet werden können. In der empirischen Sozialforschung hat dieses Problem einen weitaus geringeren Stellenwert: der Festlegung der Grundgesamtheit (Bevölkerung der Stadt XYZ), der Auswahlgesamtheit und der eventuellen Ziehung einer repräsentativen Auswahl aus der Auswahlgesamtheit folgt die Untersuchung im "Feld", d.h. die Phase der Datensammlung z.B. in Form von Interviews. Dabei ist idealiter die Relation zwischen ausgewählten Einheiten und der angestrebten Ziel- bzw. Grundgesamtheit unproblematisch, können doch in der empirischen Sozialforschung in der Vorphase der Datenerhebung mögliche Inkongruenzen zwischen Erhebungseinheit und Zielgesamtheit in der Regel korrigiert werden. Diese Chance der Korrektur ist in der historischen Sozialforschung nicht gegeben. Die auf die Definition der Zielgesamtheit folgende Bestimmung der Auswahlgesamtheit und der Erhebungseinheit ist abhängig davon, inwieweit diese Erhebungseinheiten in dem zur Verfügung stehenden historischen Quellenmaterial "wiederzufinden" sind und in welchem Maße diese Erhebungseinheiten der angestrebten Zielgesamtheit entsprechen.

Dieses Problem soll an Hand eines Forschungsbeispiels verdeutlicht werden: Gegeben sei eine Forschungsfrage, die Ursachen, Formen und Ausmaße delinquenten Verhaltens der preußischen Bevölkerung im Zeitalter der Hochindustrialisierung, also in der Zeit des Kaiserreichs, untersuchen soll. Ziel- bzw. Grundgesamtheit ist also in diesem Falle die delinquente Bevölkerung Preußens im Zeitraum 1871-1914. Die Festlegung der Erhebungseinheit könnte theoretisch in verschiedenen Varianten resultieren: Einerseits könnte als Erhebungseinheit das einzelne delinquente Individuum herangezogen werden, in gleichem Maße könnten aber auch Aggregate von Personen als Erhebungseinheit fungieren. Quellenmaterial, an Hand dessen sich entsprechende Erhebungseinheiten festlegen und für Analysen in Untersuchungen aufnehmen lassen, sind vorhanden: Würde man versuchen, das einzelne delinquente Individuum als Erhebungseinheit in Betracht zu ziehen, so hätte man auf gerichtliche Untersuchungsakten zu rekurrieren; würde man als Erhebungs-

einheit Aggregate von Personen heranzuziehen versuchen, so könnte man auf die amtlichen Zählungen der gerichtlichen Kriminalstatistik des Statistischen Reichsamts und des Reichsjustizamtes zurückgreifen. Auf diese Quellen beschränkt sich im wesentlichen das Material zur obengenannten Forschungsfrage; in dieser Beschränkung ist aber zugleich auch der Grund dafür zu suchen, daß eine adäquate Abbildung einer Zielgesamtheit "delinquente Bevölkerung Preußens in der Hochindustrialisierung" nicht zu erreichen ist, sondern nur Annäherungswerte möglich sind.

Als generelles Problem ist zusätzlich zu konstatieren, daß die genannten Quellen nicht auf die delinquente Person allgemein, sondern auf *erfaßte* delinquente Personen abstellen, was für sich allein genommen die Beantwortung der Forschungsfrage – so wie sie gestellt worden ist – schon schwierig erscheinen läßt. Von dieser allgemeineren Frage der Messung von Delinquenz einmal abgesehen, stellen sich noch weitere Probleme: Auf das Individuum als Erhebungseinheit und damit auf Fallakten der Justiz zu rekurrieren, wird mit Sicherheit dazu führen, daß nicht alle erfaßten delinquenten Personen in Preußen im genannten Zeitraum, bzw. eine repräsentative Auswahl, falls eine Stichprobe gezogen werden sollte, in die Untersuchung einbezogen werden können; zu groß wird der Verlust an Untersuchungsakten durch Kassation oder durch Kriegsverlust sein, ja es ist vorstellbar, daß Untersuchungsakten für ganze Gebietseinheiten verlorengegangen sein könnten. Im Falle der publizierten Kriminalstatistiken des Deutschen Reiches ist das Überlieferungsproblem nicht relevant, weil die publizierten Tabellen aus der Reihe der "Statistik des Deutschen Reiches" in den wissenschaftlichen und sonstigen Bibliotheken im Laufe der Zeit kaum Verluste erlitten haben; und wenn Verluste eingetreten sind, sind diese in der Regel ersetzt worden. Jedoch erreichen auch die Kriminalstatistiken nur eine Annäherung an die Zielpopulation "delinquente Bevölkerung Preußens im Zeitalter der Hochindustrialisierung". Dennoch bedeuten derartige "Defekte" im Quellenmaterial nicht unbedingt, daß nunmehr die Fragestellung geändert werden muß. Die Möglichkeit, fehlende Informationen per Hochrechnung oder per Schätzung auszugleichen, ist sicherlich in vielen Fällen gegeben; allerdings verlangen derartige Korrekturen eine genaue Kenntnis des Materials.

Wenn man auch hier einmal davon absieht, daß bei den Kriminalstatistiken *erfaßtes* delinquentes Verhalten wiedergegeben wird, so bleibt als Problem das der inhaltlichen Eingrenzung: Die Kriminalstatistiken erfassen nur Verstöße gegen Reichsgesetze, ein breites Spektrum bundesstaatlicher Strafgesetzgebung unterhalb der Ebene der Reichsgesetze wird dagegen nicht erfaßt, d.h. eine bestimmte Gruppe der erfaßten delinquenten Bevölkerung



bleibt ausgeblendet. Damit ist auch hier wiederum nur eine Annäherung an die gesuchte Zielgesamtheit erreicht.

Ähnliche Beispiele lassen sich auch zu anderen Fragen aus anderen Bereichen der historischen Sozialforschung finden. Als grundsätzliches Fazit für den Datenbeschreibungs- und Datendokumentationsbedarf in der historischen Sozialforschung auf der Ebene von Zielgesamtheiten und Erhebungseinheiten läßt sich festhalten: Erörterungen über Aussagekraft und Reichweite herangezogener Quellen müssen – über den Rahmen der traditionellen Fragen der historischen Quellenkritik hinaus – Antworten auf Fragen nach der Abbildungsqualität von Quellen im Hinblick auf die angestrebte Zielgesamtheit beinhalten. Diese Antworten müssen dokumentiert werden. Ohne Ausführung über die Relation zwischen der auf bestimmten Quellen beruhenden Erhebungseinheit und der intendierten Zielgesamtheit kann die Gültigkeit der betreffenden Untersuchung nicht beurteilt werden.

## **2.2 Indikator und Quelle in der historischen Sozialforschung**

Die Indikatorenbildung ist in der Sozialforschung – neben der Definition von Grundgesamtheit, Auswahlgesamtheit und Erhebungseinheit – ein wichtiger Bestandteil der Konzeptualisierungsphase einer empirischen Untersuchung. Ausgangspunkt dieses Vorgangs ist in der Regel eine theoretische Forschungsfrage. Zumeist ist eine derartige Forschungsfrage nicht direkt beantwortbar, d.h. sie ist nicht ohne weiteres überprüfbar, meßbar. Dazu bedarf es der Umsetzung des theoretischen Sachverhalts in beobachtbare Größen, d.h. es werden Sachverhalte (Einstellungen, Verhalten, etc.) definiert, die den theoretischen Sachverhalt indizieren. Die Antwort auf eine Frage nach dem Indikator, der diese Funktion erfüllen kann, ist in den Modellvorstellungen der empirischen Sozialforschung primär abhängig von theoretischen Vorüberlegungen und Ergebnissen empirischer Vorarbeit. In der Forschungspraxis, insbesondere in der Sekundäranalyse, ist zwar das Verhältnis zwischen zu messendem Sachverhalt und Indikator häufig interaktiv; dennoch ist immer noch dieses deduktive Modell der Ermittlung von Indikatoren stilbildend. Dieser deduktive Ansatz hat Konsequenzen für das Verhältnis von Indikator und sozialem Sachverhalt, der mit Hilfe des Indikators gemessen werden soll. Die Definition von Indikatoren zwecks Messung der in der Forschungsfrage angesprochenen Sachverhalte führt im Ergebnis dazu, daß auf der Grundlage dieser Entscheidungen und daraus resultierender Meßoperationen die für eine Untersuchung relevante Realität bestimmt, "selektiert" wird. Man könnte mithin sagen, daß das Datenmaterial einer empirischen Untersuchung abhängig von den definitorischen Eingrenzungen der

die Untersuchung durchführenden Personen ist, daß die in der Untersuchung abgebildete Realität durch einen Selektionsvorgang in der Konzeptualisierungsphase determiniert wird. In der historischen Sozialforschung stellt sich dieses Problem anders. Zwar entspricht sich im Ansatz die Indikatorenbildung in beiden Zweigen der Sozialforschung, der Tatbestand jedoch, daß der historisch arbeitende Sozialforscher an bereits vorliegendes Material gebunden ist, hat weitergehende Rückwirkungen auf die Indikatorenbildung. Zwar kann im Idealfall, d.h. bei günstiger Quellenlage, in gewissem Maße auch eine "freie" Selektion der für die Untersuchung relevanten Realität durch die Indikatorenbildung stattfinden. Üblicherweise muß man in der historischen Sozialforschung aber davon ausgehen, daß die Indikatorenbildung an das vorhandene Material gebunden ist; wenn – vereinfacht formuliert – vorhandenes Einkommen als ein "starker" und Beruf ein "schwacher" Indikator für soziale Schichtung angesehen werden, die Quellen jedoch keine Einkommensangaben enthalten, so muß Beruf als Indikator für Schichtung berücksichtigt werden. Die für eine Untersuchung relevante Abbildung der Realität liegt in der historischen Sozialforschung somit bereits vor und wird nicht erst im Rahmen der Indikatorenbildung konstituiert. Damit stellt sich für die historische Sozialforschung verschärft das Problem, ob ein Indikator einen theoretischen Sachverhalt adäquat "übersetzen" kann.

In der Forschung wird dieser Problembereich auch im Rahmen der Diskussion über die Kontextgebundenheit historischer Daten bzw. prozeß-produzierter Daten diskutiert. Historische Quellen wurden in der Regel zu anderen Zwecken geschaffen als dem, auf Fragen der Sozialforschung von heute Antworten zu bieten. Die Vermögenseinteilung eines städtischen Steuerverzeichnisses des ancien régime genügt sicherlich den fiskalischen Interessen der damaligen Stadtschreiber und ihrer Auftraggeber; ob diese Einteilung aber auch als adäquat angesehen werden kann, einen Sachverhalt wie Stand oder soziale Schichtung zu indizieren, steht zur Diskussion. Dies hat zur Konsequenz, daß die Kontextabhängigkeit von historischem Quellenmaterial diskutiert und im Hinblick auf den im Rahmen einer Forschungsfrage gemeinten theoretischen Sachverhalt evaluiert werden muß. Die Diskussion und die Evaluation der Quelle müssen zum "Verständnis" von Daten und zur Analyse "weitertransportiert" werden, d.h., sie müssen dokumentiert werden. Diese Informationen ermöglichen eine Beurteilung der Qualität der Primärforschung, lassen aber auch zu, daß Daten auch im Rahmen von sekundäranalytischen Nutzungen verwendet werden können.

Aus der geschichtswissenschaftlichen Diskussion quellenkritischer Fragen, vor allem aber aus den Forschungen über die Verwendbarkeit zeitgenössischer prozeß-produzierter Daten lassen sich einige generelle Kriterien

ableiten, an Hand derer Indikatorenprobleme evaluiert werden können und die zum Verständnis der Daten und der Analyse einer Untersuchung dokumentiert werden müssen. Zu diesen Kontextmerkmalen gehören:

1. Merkmale, die die faktischen und normativen Entstehungsbedingungen und die "Geschichte" von Quellen betreffen (z.B. Aufgabenprogramme, aktenproduzierende Organisationen, Zweck der Aktenanlage, etc.).
2. Merkmale, die das formelle und informelle Verhalten des Personals quellen- bzw. aktenproduzierender Organisationen berühren (spezielle Formen der Aktenführung, "Codes" in Akten, all diejenigen Merkmale, die als "Realität eigener Art" von Akten bezeichnet werden)
3. Merkmale, die – im Falle personenbezogener Quellen bzw. Massenakten – auf spezifische Verhaltensweisen der erfaßten Personen gegenüber der aktenführenden Instanz rekurrieren (Bereitschaft, bestimmte Informationen weiterzugeben, in Relation zu den Absichten der aktenführenden Instanzen, etc.).

Ad 1: Akten als Hinterlassenschaft des Aufgabenvolumens staatlicher Organisationen sind hierfür ein klassisches Beispiel. Aktenvermerke und Berichte bayrischer Provinzbeamter über den "sittlichen" Verfall der Landbevölkerung in der ersten Hälfte des 19. Jahrhunderts sind dazu verwendet worden, Aussagen über den Wandel der Sexualmoral in dieser Zeit zu formulieren. Inwieweit solche Berichte derartige Schlüsse zulassen und nicht eher nur die Sichtweise der kontrollierenden Instanzen indizieren, ist von denjenigen Wissenschaftlern zu beantworten, die diese Forschungsfrage gestellt haben. Die Frage, ob diese quellen als Indikatoren für andere Sachverhalte Gültigkeit besitzen, kann nur dann beantwortet werden, wenn die Informationen über die Entstehungsbedingungen des Materials weitere Aussagen zulassen.

Ad 2: Eine weitere Dimension des Quellenkontextes im Rahmen der Diskussion über die Indikatorenqualität von Informationen aus Quellen betrifft das Verhalten, den Verwaltungsstil derjenigen Individuen, die die Quelle erstellt haben. Auch hierzu ein Beispiel: Aus einer Untersuchung über den Widerstand gegen das NS-Regime in Düsseldorf wissen wir z.B. über die Indikatorenqualität von Vernehmungsprotokollen der Gestapo, daß die Akten häufig Sprachcodes enthalten, daß Sachverhalte systematisch ausgeblendet werden, indem sie in Handakten "versteckt" werden, oder aber indem für sie das Prinzip der Schriftlichkeit außer Kraft gesetzt wird. Derartige Tatbestände müssen wiederum sorgfältig diskutiert und dokumentiert werden, um die Indikatorenqualität des verwendeten Materials beurteilen zu können.

Ad 3: Antworten auf die Frage, ob das, was ursprünglich gemessen werden sollte, auch tatsächlich im Kontext der Quelle ermittelt wurde, sind nicht allein davon abhängig, wie sich Mitglieder aktenführender Instanzen verhielten. Das Verhalten auch derjenigen, über die Informationen erhoben wurden, spielt eine Rolle. Auch hierzu ein Beispiel: Besitzangaben in Quellen sind beliebte Informationen, um bestimmte Aspekte des sozio-ökonomischen Status von Personen zu bestimmen. Die Bereitschaft zur Offenlegung von Besitzangaben ist unterschiedlich. Besitzangaben, die dazu gedacht sind, die Höhe von steuerlichen Abgaben zu bestimmen, sind sicherlich anders einzuschätzen als Vermögensangaben, die die Grundlage für Wahlberechtigungen bilden. Wiederum einen anderen Stellenwert haben die Besitzangaben, mit denen Armenverwaltungen die Höhe von Zuteilungen begründen.

### **3. Empfehlungen für eine Verbesserung der Datenbeschreibung und Datendokumentation in der historischen Sozialforschung**

#### **3.1 Vorbemerkung zu den Empfehlungen**

Im vorausgegangenen Teil ist grundsätzlich der Datenbeschreibungs- und Datendokumentationsbedarf formuliert worden, der bei Untersuchungen der historischen Sozialforschung akut wird. Derartige Datenbeschreibungen und Datendokumentationen sind in zweierlei Hinsicht notwendig: Einerseits ist zur Bewertung und Diskussion von Ergebnissen der Forschung ein "Verständnis" der Daten einer Untersuchung erforderlich. Darüber hinaus erleichtert eine Optimierung derartiger Datenkennungen die weitere Nutzung maschinenlesbarer historischer Forschungsdaten im Rahmen von Sekundäranalysen. Allgemein sollen die im folgenden vorgeschlagenen Richtlinien dazu dienen, den bisher unzureichenden Zustand von Datenbeschreibung und Datendokumentation in der historischen Sozialforschung zu verbessern. Denn: Verbesserte Datenkennungen können helfen, die Archivierung maschinenlesbarer historischer Forschungsdaten zu erleichtern und angemessen archivierte und dokumentierte Daten für Sekundäranalysen zur Verfügung zu stellen. Die vorliegenden Empfehlungen stellen die Erfahrungen mehrerer Jahre intensiver Beschäftigung mit maschinenlesbaren historischen Forschungsdaten dar und verstehen sich als Versuch, sowohl zu einem besseren "Verständnis" von Forschung als auch zu einem leichteren Zugang zu Forschungsdaten zwecks Sekundäranalysen zu gelangen, so wie auch die Verwendung von Forschung für Forschung in einem praktischen Sinne zu verbessern.

Die Strukturierung der Beschreibungs- und Dokumentationsitems in der vorliegenden Empfehlung folgt dem bereits genannten Instrument zur Be-

schreibung und Dokumentation von Umfragedaten. Vor einigen Jahren ist dieses Instrument auf Anregung des Zentrums für historische Sozialforschung um Bestandteile erweitert worden, die es ermöglichen, auch maschinenlesbare historische bzw. prozeß-produzierte Daten zu beschreiben und zu dokumentieren. Diese Standardisierung von Datenbeschreibung und -dokumentation auch für historische Daten soll nunmehr einen Datenaustausch auch über nationale Grenzen hinweg ohne größere Probleme erlauben.

### **3.2 Empfehlungen zur Beschreibung von Grundgesamtheit und Erhebungseinheiten**

Ein Instrument zur Beschreibung und Dokumentation maschinenlesbarer historischer Forschungsdaten muß notwendigerweise die Bereiche abdecken, die in den grundsätzlichen Ausführungen zu Beginn dieses Beitrags als beschreibungs- und dokumentationsbedürftig für das "Verständnis" von Daten aus der Sozialforschung insgesamt definiert worden sind. Dazu gehören Aussagen über

- A. Die Ziel- bzw. Grundgesamtheit, auf die die Aussagen einer Untersuchung bezogen sind, und die evtl. Auswahl bzw. Stichprobe, die aus einer derartigen Gesamtheit gezogen worden ist.
- B. Die Merkmalsdimensionen (Variablen) einer Untersuchung / Kriterien zur Beurteilung der Indikatorenqualität der herangezogenen Quelleninformationen.

Den Informationen über die Grund- bzw. Zielgesamtheit einer Studie bzw. über die daraus gezogene Auswahl bzw. Stichprobe sind Informationen über Art und Herkunft der verwendeten Daten vorangestellt. Diese Kategorie, die in dem ursprünglich für die Beschreibung von Umfragedaten konzipierten Instrument nicht vorhanden war, ist für die historische Sozialforschung von zentraler Bedeutung. Neben der Auflistung der verwendeten Quellen, spezifiziert nach den verschiedenen eventuell verwendeten Quellenarten und ihrer Herkunft und Aufbewahrung, sollte an dieser Stelle vermerkt werden, inwieweit verschiedene Quellen zu einem gemeinsamen Datensatz zusammengefügt worden sind.

Die Auflistung von Art und Herkunft des verwendeten Quellenmaterials unterschieden nach unveröffentlichten und veröffentlichten Aktenmaterialien, nach veröffentlichten und unveröffentlichten Zensusmaterialien, nach "Sekundärmaterialien" oder anderen Materialien, ist an sich nichts Neues, insoweit es die Arbeit mit historischem Datenmaterial betrifft. Diese Regel ist jedoch hier noch einmal aufgeführt, um sie ausdrücklich in den Zusammenhang der Datenbeschreibung und Datendokumentation für die historische

Sozialforschung einzufügen. Zugleich soll sie aber dahingehend erweitert werden, daß die in der historischen Sozialforschung nicht unübliche Praxis der Kombination von Quellen zwecks Einfügung fehlender Informationen "am Ort", d.h. bei der Auflistung der herangezogenen Quellen, dokumentiert wird. Dieser Basisinformation folgt der Abschnitt der Datenbeschreibung und Datendokumentation, dessen Inhalt Aussagen über Ziel- und Grundgesamtheit einer Untersuchung und die daraus evtl. gezogene Auswahl bzw. Stichprobe sind.

Die Datenbeschreibung und -dokumentation beschränkt sich in diesem Abschnitt jedoch nicht auf diese beiden Fragen, sondern bezieht eine Reihe weiterer Probleme mit ein. Vorrangig in diesem Beschreibungskomplex ist die Definition der Ziel- bzw. Grundgesamtheit. Damit ist der Rahmen abgesteckt, innerhalb dessen sich das Forschungsinteresse der betreffenden Untersuchung lokalisieren läßt. Mit der Ziel- bzw. Grundgesamtheit werden in der Regel Auswahlgesamtheit und Erhebungseinheit definiert. Zur Berichterstattungspflicht des Wissenschaftlers in der historischen Sozialforschung gehört es, exakte Explikationen der Auswahlgesamtheit und der Erhebungseinheiten zu geben. Inwieweit diese dann die angestrebte Ziel- bzw. Grundgesamtheit abdecken, ist eine Frage, die der Forscher im Rahmen seiner quellenkritischen Überlegungen zu entscheiden hat. Die Ergebnisse dieser Überlegungen sind "mitteilungspflichtig", lassen sie doch Schlüsse darüber zu, inwieweit die aus den Quellen definierten Erhebungseinheiten adäquat (im Sinne der Repräsentativität) die angestrebte Zielgesamtheit wiedergeben und für welche anderen Fragestellung sie eventuell eingesetzt werden könnten.

Eine Kennzeichnung der Relation zwischen angestrebter Zielgesamtheit und den durch die vorhandenen Quellen festgelegten Erhebungseinheiten erlaubt eine Beurteilung der Reichweite (der intendierten und bei günstiger Quellenlage auch erreichten) Forschungsaussage und ist für eine weitere Verwendung der jeweils beschriebenen und dokumentierten Daten unerlässlich. Es ist durchaus vorstellbar, daß Erhebungseinheiten im Rahmen sekundär-analytischer Nutzungen auch für anderen Ziel- bzw. Grundgesamtheiten repräsentativ sein können als für die, die im Rahmen der Originalerhebung definiert worden sind. Die Notwendigkeit, Korrespondenz von angestrebter Ziel- bzw. Grundgesamtheit und den Erhebungseinheiten zu explizieren, ist auch dann gegeben, wenn keine Stichprobe gezogen, sondern eine Total- bzw. Vollerhebung vorgenommen worden ist. Da die Erhebungseinheiten durch das vorliegende Quellenmaterial präjudiziert werden, ist dieser Problembereich grundsätzlicher Natur, d.h. er berührt die Quellenproblematik von Untersuchungen der historischen Sozialforschung.

Ist jedoch aus dem vorliegenden Quellenmaterial eine Stichprobe gezogen worden, so ist die Kennzeichnung des angewandten Stichprobenverfahrens und der Vorgehensweise bei der Stichprobenziehung Teil der Berichtspflicht des Forschers. Entsprechende Angaben erlauben es, die Repräsentativität der untersuchten Teilmengen einer Untersuchung im Hinblick auf die angestrebte Ziel- bzw. Grundgesamtheit zu beurteilen. Bestandteil der Kennzeichnung der Auswahl sollten Angaben sein, die eine Beurteilung der Entsprechung zwischen Merkmalen der ausgewählten Einheiten und der Zielgesamtheit zulassen. Diese Frage kann eventuell auch auf das Problem der Entsprechung von Zielgesamtheit und den durch das vorliegende Quellenmaterial bestimmten Erhebungseinheiten zurückführen, andererseits aber auch ein Verweis auf Probleme der Vorgehensweise bei der Stichprobenziehung sein, deren Kennzeichnung nicht nur für das "Verständnis" der Daten, sondern erneut auch für sekundäranalytische Nutzungen von Daten von Bedeutung ist.

In der Diskussion über die grundsätzlichen Probleme der Abbildung einer bestimmten Grund- bzw. Zielgesamtheit durch bestimmte Quellen wurde auf die Überlieferungsproblematik als einem möglichen Faktor verwiesen, der zu einer verzerrten Abbildung der Grundgesamtheit durch die Quellen führen kann. Um derartige Verzerrungen zu erfassen, wurden auf der Grundlage der vom Zentrum entwickelten Datenbeschreibungs- und Datendokumentationsmerkmale besondere Kennungen der Quellenüberlieferung aufgenommen. Damit sollte es in Zukunft möglich sein, Verzerrungen im Hinblick auf die zu bildende Grundgesamtheit zu beurteilen.

Um die Überlieferungsproblematik der verwendeten Quellen zu verdeutlichen und zu kennzeichnen, sollen die im folgenden aufgeführten Fragen nach Zustand und Erhaltung von Quellen beantwortet werden. Archivare, die in ihrer alltäglichen Arbeit weitaus intensiver mit den Problemen der Quellenüberlieferung befaßt sind, werden die hier vorgeschlagenen Kategorien sicherlich als nicht ausreichend empfinden. Demgegenüber muß aber darauf verwiesen werden, daß die vorgeschlagenen Kategorien Kompromisse zwischen einem wünschbaren Optimum an Informationen über Quellenbestände und den Einschränkungen, die aus der Standardisierung eines Instruments resultieren, darstellen. Ganz wesentlich ist auch, daß hier eine offene Frage gestellt wird, welche das Ausmaß der Beantwortung dem Einzelfall überläßt. Die Überlieferungsproblematik einer Quelle sollte im einzelnen entlang den folgenden Merkmalen gekennzeichnet werden. Abhängig vom "Zustand" der verwendeten Quelle können diese Kennzeichnungen durchaus mehrere der genannten Merkmale betreffen. Angaben sollten darüber gemacht werden:

- ob die verwendete Quelle vollständig erhalten ist,
- ob die verwendete Quelle nur teilweise erhalten ist, d.h. ob sie eventuell gemäß bestimmten, vielleicht auch nachvollziehbaren Kriterien vom Archivpersonal teilkassiert worden ist,
- ob die Quelle teilkassiert worden ist, ohne daß bestimmte Kassationskriterien angebbbar sind,
- oder ob die Quelle vielleicht teilzerstört ist, weil der Bestand aufgrund äußerer Einwirkungen (Feuer, Krieg, sonstige physische Beeinträchtigungen) beschädigt ist.

Neben der Bewertung der Fragen, in welchem Maße die verwendete Quelle erhalten ist, ist es auch wichtig, Hinweise auf die interne Organisation des Quellenmaterials zu geben. Die Abbildung einer Grundgesamtheit in den Quellen ist auch davon abhängig, wie sichtbar Informationen in Quellen sind. Dazu sind Angaben notwendig

- über die Form der Aktenhaltung beim Quellenproduzenten, soweit diese noch aus der Quelle sichtbar ist,
- oder aber über die Form und das Ausmaß der Umorganisation der Quelle im Archiv oder bereits bei der Stelle, die die Quellen produziert hat.

Die Notwendigkeit, die Abbildung einer Zielgesamtheit durch eine Quelle einschätzen zu müssen, verlangt über die Informationen über das Ausmaß und Art der Überlieferung hinaus auch Angaben über die Dokumentation und den Zugang zur Quelle. Eine unzulängliche Quelldokumentation, ein unvollständiges Findbuch, eine nicht gewährleistete vollständige Zugänglichkeit zur Quelle können durchaus in Abbildungsfehlern resultieren. Deshalb sollte die Datenbeschreibung/Datendokumentation Informationen darüber enthalten, ob die verwendete Quelle

- vollständig erschlossen und zugänglich ist,
- ob sie eventuell nur teilerschlossen, jedoch prinzipiell zugänglich ist,
- und ob sie vielleicht sogar noch völlig unerschlossen ist.

### **3.3 Beschreibung der Merkmalsdimensionen (Variablen)/Kriterien zur Beurteilung der Indikatorenqualität der herangezogenen Quelleninformationen**

In der grundsätzlichen Diskussion der Datenbeschreibungs- und Datendokumentationsprobleme der historischen Sozialforschung sind bereits die Probleme angesprochen worden, die bei einem dem "Verständnis" der Daten der historischen Sozialforschung adäquaten Beschreibungs- und Dokumentationsinstrument berücksichtigt werden müssen. Zwei Problembereiche, die diskutiert und dokumentiert werden sollen, wurden angesprochen:



- Die Beschreibung von Kontextmerkmalen der Quelle, um die Qualität der verwendeten Indikatoren evaluieren zu können.
- Der erweiterte Codeplan, um die zeitgebundene Bedeutung von Variablen zu kennzeichnen und festzuhalten.

Die Beschreibung des Kontextes von Quellen sollte in erster Linie folgende Merkmale berücksichtigen:

- Formelle Entstehungsbedingungen von Quellen, insbesondere Gesetze, Verwaltungsanordnungen, Geschäftsordnungen, etc. Die entsprechenden Regelungen sollten aufgeführt werden, vor allem sollten auch Operationalisierungen auf administrativer Ebene, z.B. die Umsetzung von Gesetzen in Ausführungsverordnungen dokumentiert werden.
- Charakteristika der Informationsverarbeitungsstrategien von Quellenproduzenten. An dieser Stelle sollten en détail diejenigen festgestellten Verfahren, Routinen, Schreibweisen, Auslassungsstrategien, etc. des oder derjenigen festgehalten werden, die die betreffende Quelle erstellt und bearbeitet haben.
- Daneben sind, soweit feststellbar, auch diejenigen Informationen eines Quellenkontextes zu berücksichtigen, die auf die Möglichkeit rekurrieren, daß Quellen auch von denen beeinflusst sein können, auf die sich die Inhalte der Quellen beziehen. Demententsprechend sollten die Ergebnisse der quellenkritischen Diskussion, die derartige Fragen anschnitten, hier dokumentiert werden.

Der erweiterte Codeplan ist für die historische Sozialforschung unerlässlich, um den Beschreibungs- und Dokumentationsbedarf auf der Variablenebene abzudecken. In der empirischen Sozialforschung sind entsprechende Erfordernisse – trotz bestehender Probleme – bislang noch nicht aufgenommen worden.

An dieser Stelle sollen nur die Informationen formuliert werden, die geeignet erscheinen, eine Variable aus dem Bereich der historischen Sozialforschung adäquat zu "verstehen":

Codebuchinformation pro Variable

1. Nr. der Variable
2. Bezeichnung bzw. kompletter Text der Variable
3. Bezeichnung bzw. Etikett und Wertebereich der Variablenausprägungen
4. Position der Variable auf dem Datenträger (z.B. Kartenart, falls Lochkarte als Datenträger verwendet wurde, Spaltenkennungen)
5. Randauszählung
6. Variablenkommentar
7. Quellenangaben.

Während die Informationen 1 - 5 zu den üblichen Codebuchstandards der empirischen Sozialforschung gehören, gehen die beiden übrigen Informationen über die bis dato in diesem Bereich der Sozialforschung noch gültigen Beschreibungsstandards hinaus. Sie erfüllen den zusätzlichen Informationsbedarf der historischen Sozialforschung. Ein derartiger Variablenkommentar kann durchaus aus einer längeren Textstelle bestehen. Bei Angaben zu Aggregatdaten sind auch Informationen über das Meßniveau diejenigen amtlichen Zählungen einzubeziehen, auf deren Grundlage diese Daten erstellt worden sind. Ohne diese zusätzlichen Angaben sind Aggregatdaten nicht "verständlich".

Die Quelleninformation pro Variable hat Kontrollfunktion, sie dient dazu, bei Datenbereinigungen die wiedergegebenen Variablen evtl. anhand der Quelle zu überprüfen.

#### **4. Kurzfassung der Empfehlungen für eine Verbesserung der Datenbeschreibung und Datendokumentation in der historischen Sozialforschung**

- A) Beschreibung von Ziel- bzw. Grundgesamtheit, von Erhebungs- bzw. Beobachtungseinheiten und eventueller Stichproben aus der Quelle
- A1) Definition der Ziel- bzw. Grundgesamtheit:

Auf welche Ziel- bzw. Grundgesamtheit bezieht sich die Fragestellung Ihrer Untersuchung? Beispiele möglicher Antworten: Ziel- bzw. Grundgesamtheit ist die Bevölkerung der Stadt Osnabrück in der zweiten Hälfte des 19. Jahrhunderts; Ziel- bzw. Grundgesamtheit ist der Grundbesitz in der Stadt Köln den Jahren 1850 bis 1910 etc.

- A2) Explikation der Erhebungs- bzw. Beobachtungseinheiten

Anhand welcher (in den Quellen erfaßter) Einheiten haben Sie versucht, Aussagen über Ihre Ziel- bzw. Grundgesamtheit zu machen? Mögliche Antworten dazu im Rahmen der o.g. Beispiele: Erhebungs- bzw. Auswahlinheit sind alle in den Adreßbüchern der Stadt aufgeführten Haushalte, oder aber alle in den Katasterverzeichnissen der Stadt Köln vorgenommenen Eintragungen über Besitzverhältnisse.

- A3) "Repräsentativität" der Erhebungseinheiten im Hinblick auf die abzubildende Zielgesamtheit:

Hier ist die Frage zu beantworten, ob und wie die Erhebungseinheiten (die oftmals durch das vorliegende Quellenmaterial bedingt sind) die Zielgesamtheit abbilden. Um weiterhin im Rahmen eines unserer Beispiele zu argumentieren: Bilden Adreßbücher der Stadt Osnabrück deren Sozialstruktur

im gegebenen Zeitraum adäquat ab, oder muß auch in diesem Fall konstatiert werden, daß, wie häufig bei diesem Quellentyp zu beobachten, Unterschichtenangehörige nicht erfaßt werden? Wie groß schätzen Sie diese Dunkelziffer?

A4) Angaben über das verwendete Stichprobenverfahren und über die Vorgehensweise bei der Ziehung der Stichprobe:

Falls Sie eine Stichprobe aus Ihrer Quelle gezogen haben, beschreiben Sie die Vorgehensweise bei der Ziehung der Stichprobe. Die Antwort auf diese Frage sollte überaus detailliert Aufschlüsse über die verschiedenen Schritte einer Stichprobenziehung und über die Qualität der gezogenen Stichprobe zulassen. Als Beispiel einer derartigen Beschreibung (einer komplizierten Vorgehensweise) sei hier der entsprechende Abschnitt aus der bereits genannten Untersuchung über den Widerstand in Düsseldorf gegen das NS-Regime angeführt:

“Die ‘Personalakten’ der Gestapoleitstelle Düsseldorf waren nach dem Stand vom 1. Mai 1977 zu über zwei Drittel in mehreren Karteien systematisch erschlossen. In einer ‘Ortskartei Düsseldorf’ waren ca. 5000 Karteikarten über Personen gesammelt, die zwischen 1933 und 1945 in Düsseldorf ‘staatspolitisch in Erscheinung getreten’ sind und über die die Gestapo Düsseldorf eine (oder mehrere) ‘Personalakten’ angelegt hatte. Diese Ortskartei ist nach 52 ‘Hauptschlagwörtern’ in Anlehnung an den Geschäftsverteilungsplan des Geheimen Staatspolizeiamtes in Berlin vom 1.7.1939 untergliedert. 41 Schlagwörter wurden in die Erhebung einbezogen, elf blieben dagegen un bearbeitet. Als Grundgesamtheit, aus der die Auswahl gezogen werden sollte, wurden schließlich die ca. 3.770 Karteikarten in den 41 relevanten Schlagwörtern der ‘Ortskartei Düsseldorf’ definiert. Eine (personenbezogene) Karteikarte, auf der die Nummer(n) der in Frage kommenden Gestapoakten verzeichnet ist (sind), bildete die Auswahlinheit, während eine Person die Analyseinheit darstellt. Die 41 ‘Hauptschlagwörter’ wurden in vier Kategorien gruppiert und für jede Kategorie Quoten festgelegt; als Stichprobenumfang ergab sich eine Fallzahl von 825 Personen. Im Zuge einer ersten Auswahlstufe wurde aus den ‘Hauptschlagwörtern’ mit mehr als 90 Karteikarten eine Wahrscheinlichkeitsauswahl von 10 % gezogen (Ziehen jeder 10ten Karteikarte mit Zufallsstart). Aus Schlagwörtern mit weniger als 90 Karteikarten wurden die Karteikarten ‘vollständig’ erhoben; allerdings blieben sog. ‘Sammelsachakten’ zunächst unberücksichtigt, so daß auch hier praktisch systematische Auswahl gezogen wurden. Durch diese Verfahren gelang es, das breite Spektrum unterschiedlicher Tatbestände in der Auswahl zu erfassen. Die umfangreichen Tatbestandsgruppen, z.B. im ‘Hauptschlagwort’ ‘Heimtücke’, sind in der Auswahl zugunsten der Tatbestände mit geringen

Fallzahlen unterrepräsentiert. Dies gilt in besonderem Maße für die KPD: Die Karteikarten in diesem Schlagwort umfassen ca. 28 % aller relevanten Karteikarten, in unserem Sample stammen dagegen nur knapp 14 % der Fälle aus diesem Schlagwort, das somit zu 100 % unterrepräsentiert ist" (Reinhard Mann, Protest und Kontrolle im Dritten Reich. Nationalsozialistische Herrschaft im Alltag einer rheinischen Großstadt, Köln und Wien: Böhlau Verlag, erscheint 1985).

A5) Angaben über das "Überleben" der Quelle:

Hier sind die Angaben darüber zu machen, ob die Quelle eventuell vollständig erhalten ist, oder ob sie eventuell teilkassiert oder zerstört worden ist. Soweit Teilkassationen vorgenommen worden sind, sind Ihnen die Kassationskriterien bekannt (geworden)? Ist vielleicht systematisch ein Teilbestand kassiert worden, oder aber streuen sich die Kassationen über den gesamten Bestand? Kann man etwas über den Effekt sagen, den eventuelle Teilerstörungen hatten? Falls Sie eine Stichprobe gezogen haben, wie haben Sie derartige Kassationen/Zerstörungen berücksichtigt?

A6) Angaben über die Ordnung der Quelle:

Hier sind Fragen zu beantworten, die die Ordnung des Quellenmaterials betreffen. Falls das Quellenmaterial Ihrer Untersuchung personenbezogene Massenakten sind, wurden die Quellen alphabetisch, chronologisch oder geographisch oder in Kombinationen hiervon sortiert? Sind diese Sortierungen in der Vergangenheit eventuell verändert worden? Kennen Sie z.B. Anweisungen für die Aktenführung, die von der aktenhaltenden Instanz ausgegeben worden sind? Wenn ja, zitieren Sie diese Anweisungen! Sind einzelne Akten des Bestandes im Zeitverlauf ergänzt worden, nicht jedoch alle Akten insgesamt? Ist die Quelle nach der Abgabe an das Archiv in der Weise verändert worden, daß z.B. eine ursprünglich vorhandene Sortierung zugunsten einer anderen Sortierung verändert worden ist.? Welche Konsequenzen hat dies für Ihre Stichprobenbeziehung gehabt?

A7) Angaben über die Erschließung und die Zugänglichkeit einer Quelle:

Mit welchen Mitteln und in welchem Umfang war die von Ihnen herangezogene Quelle erschlossen? Existierte ein Findbuch, war die Quelle per Karteikarte teilweise oder komplett erschlossen? Gab es noch den ursprünglichen Aktenplan? Existieren überhaupt Informationen über die Sortierungskriterien der Quellen?

B) Beschreibungen der Merkmalsdimensionen, Kriterien zur Bewertung von Indikatoren, Variablenkommentare

B1) Angaben zu den formellen Entstehungsbedingungen von Quellen:

Mit welchen Intentionen wurde die für Ihre Untersuchung relevante Quelle ursprünglich geführt? Machen Sie hier bitte Angaben über den formellen Auftrag, der (in der Regel) zur Entstehung einer Quelle geführt hat (im Falle von Massenakten der öffentlichen Verwaltung) oder aber die Absicht, die hinter der Entstehung anderer Quellen (wie z.B. Adreßbücher) stand.

**B2) Angaben über das Verhalten von Quellenproduzenten:**

Ob anhand vorhandener Angaben in den Quellen das gemessen wird, was gemessen werden sollte, ist auch abhängig von verschiedenen Faktoren. Dazu sind Angaben über die faktischen Verfahrensweisen bei der Quellenerstellung notwendig. Ähnlich den Codes, die heute in Zeugnissen über ausscheidende Mitarbeiter einer Firma üblich sind, können auch historische Quellen spezifische Verfahrensweisen reflektieren, die beschrieben werden müssen, um die Indikatorfunktionen der Quelle beurteilen zu können.

**B3) Angaben über mögliches Verhalten der "Objekte" von Quellen:**

Die Frage, ob die Personen, über die Informationen in der betreffenden Quelle erhoben worden sind, Einfluß auf diese Informationen hatten, ist in diesem Kontext ebenfalls zu beantworten. Gab es ein Interesse dieser Personen, Informationen eventuell zu selektieren, z.B. die Höhe persönlichen Besitzes je nach Kontext unterschiedlich anzugeben? Konnten Sie derartige kontextabhängige Verzerrungen überprüfen; war es Ihnen möglich, eventuell Korrekturen vorzunehmen?

**B4) Variablenkommentar im Codebuch:**

Bitte, kommentieren Sie die Bezeichnung der Variablen und der Variablenausprägung so detailliert wie möglich. Geben Sie Quelleninformationen pro Variable. Listen Sie die folgenden Informationen pro Quelle auf:

1. Nr. der Variable, 2. Bezeichnung bzw. kompletter Text der Variable, 3. Bezeichnung bzw. Etikett und Wertebereich der Variablenausprägung, 4. Position der Variable auf dem Datenträger, 5. Randauszählung, 6. Variablenkommentar, 7. Quellenangabe.

Anbei noch ein Beispiel für einen Variablenkommentar. In diesem Kommentar wird die Variable "Politische Position der Abgeordneten" der französischen Nationalversammlung erläutert. Das Beispiel entstammt einer laufenden Untersuchung:

"In dieser Variablen wurden die im Primärmaterial enthaltenen Kennzeichnungen der politischen Position des Abgeordneten erfaßt. Diese Angaben sind vage und verwenden keine einheitliche Terminologie. Als Hinweis auf die politische Position des Abgeordneten führen die Autoren zuweilen Urteile zeitgenössischer Beobachter an, Robert/Cougny werten

zusätzlich Abstimmung aus (nach Raginel). Gelegentlich bleibt unklar, ob sich die Einordnung auf die Position des Abgeordneten während der Julimonarchie oder während der Revolution von 1848/49 bezieht. Problematisch bleibt insbesondere die Untergliederung des rechten Flügels in Legitimisten, Orleanisten, Bonapartisten und 'unspezifische Rechte'. In den häufigen Zweifelsfällen wurde der Abgeordnete der Kategorie 'Rechte' zugeordnet. Ähnlich zweifelhaft ist die Unterscheidung von 'Gemäßigten Linken' und 'Gemäßigten Republikanern'. Auf dem linken Flügel ist die Einordnung eindeutiger und zuverlässiger. Zu einer exakten Zuordnung reichen die Angaben jedoch nicht aus. Dies leisten die aus der seriellen Analyse des Abstimmungsverhaltens gewonnenen Indizes.

In der Variablen PPI wurde die (zeitlich) erste politische Position erfaßt, die Variable PP2 erfaßt die Zielposition derjenigen Abgeordneten, die ihren Standort gewechselt hatten. Auch über diese Positionswechsel informiert zuverlässiger die serielle Analyse des Abstimmungsverhaltens." (Heinrich Best, *Die Abgeordneten der Assemblée Nationale Constituante 1848*, Köln: Zentrum für historische Sozialforschung 1983 (= Datenhandbücher für die historische Sozialforschung, Bd. 2), S. 37 f.)

## Anhang

Probleme der Datenbeschreibung und Datendokumentation sind in sozialwissenschaftlichen Lehrbüchern bislang noch unterbelichtet. Ein Vorschlag für einen Katalog sozialwissenschaftlicher Datenbeschreibungselemente wurde Anfang der 80er Jahre vorgelegt:

- *Alice Robbin*: Technical Guidelines for Preparing and Documenting Data, In: *Robert Boruch et al. (eds.): Reanalyzing Programm Evaluations*, San Francisco: Jossey Bass, 1981, S. 84-143.

Die Diskussion um die spezifische sozialwissenschaftliche Nutzung historischer und prozeß-produzierter Daten ist vor allem in den folgenden Arbeiten und Sammelbänden geführt worden:

- *Paul J. Müller (Hrsg.): Die Analyse prozeß-produzierter Daten*, Stuttgart: Klett-Cotta, 1977.
- *Wolfgang Bick, Paul J. Müller*: Probleme der Nutzung prozeß-produzierter Daten, Bonn: Bundesministerium für Forschung und Technologie, 1982.
- *Wolfgang Bick, Reinhard Mann, Paul J. Müller*: Sozialforschung und Verwaltungsdaten, Stuttgart: Klett-Cotta, 1984.



## Historical Databases and the Researcher

The title for this paper was chosen in haste. I only realised the significance of this fact later when I started to write down some ideas on the subject. I found myself asking two questions, 'what is an historical database, and who is the researcher?'. Decisions had to be made. Following some thought on the matter the conclusion was reached that an historical database should include any item of information which an historian perceives as being of use for the construction of historical fact and theory. Therefore, a database may have within it archives as housed by the P.R.O and local record offices; transcripts and research workings such as those in the library of the Society of Genealogists; manuscripts and published works of historical interest. On the other hand, the researcher can be seen to be anyone who may potentially use or add to this collection. However, these definitions are unfortunately too diverse for the context of this paper, therefore a restriction has had to be introduced. It was decided to narrow the range by including in the historical database only documentary sources available in machine-readable form<sup>1</sup>. The definition of researcher however remains unchanged, yet it is important to note that anyone should be able to use such a database since this qualification would exclude local data exchange networks, such as that operated by the users of Quarry Bank 1851, a computer software package for the analysis of census material<sup>2</sup>.

The inclusion of the words 'computer' and 'historian' in the same sentence may to many appear incongruous. Historians, not entirely by accident, have often been cast as a traditional, conservative sort, not the kind of people to be dramatically influenced by recent technological inventions. Unfortunately it appears that this view is not completely ill-founded. Ten years ago a report assessing the impact of computers on the study of social science within universities concluded that history was last, a long way down the

---

<sup>1</sup> However it has been suggested that published material in machine-readable form obtained from automated typesetting machines should also be included in such a database. See *M. Thaller's* contribution in this volume.

<sup>2</sup> T. C. Lewis and G. S. Nunn: *Quarry Bank 1851: Census Data Retrieval and its application in schools*, Heinemann Computers in Education, London, 1983. Longman also plan a data exchange for users of their computer package *Census Analysis*.



field, behind the other disciplines<sup>3</sup>. Eighty-five percent of all social science departments used a computer, yet only 26 percent of history departments did so, the nearest rival was politics with 77 percent of departments using computers. In 'real' Terms the figure of 26 percent meant that just eight history departments in England indicated that they had made use of a computer. Of individuals, only 5 percent of historians had used a machine, whereas 35 percent of all social scientists had. Education was the next lowest subject in which 24 percent of individuals had become involved with computing<sup>4</sup>. No doubt, in the period since this report the number of historians using computers has markedly increased. However, if a comparable report were to be conducted now, it might well reveal that the gulf between historians and their fellow social scientists had widened. Indeed, although the overall number of computer using historians has grown, with regard to the 'state-of-the-art', in some respects the situation may well have deteriorated. This point will be clarified later on. Of technological innovations affecting historical research the computer is to be found a long way down the pecking-order, behind both the development of the printing press and the invention of the photocopier<sup>5</sup>. As such, the greatest use to which historians put the computer is undoubtedly word-processing.

So far a dark and gloomy background to computerised historical research has been painted. To this it is hoped that a lighter image will be added, yet when viewing the picture as a whole it must be realised that it is the contrast between the two that probably forms the most important and striking aspect. One brighter feature is that there already exists a national data archive, situated at the University of Essex; set up to collect, house and distribute machine-readable datasets relevant to research in the social sciences<sup>6</sup>. The bulk of the historical datasets available from the archive are based on the nineteenth-century census returns, notably those from the national two percent sample of 1851, deposited by Professor Michael Anderson of Edinburgh University, but the collection also includes files of aggregated parish

---

<sup>3</sup> *Computing and the Social Sciences: A Report to the SSRC*, Social Science Research Council, London, 1973. See especially Appendix II, pp. 23-25.

<sup>4</sup> One must not, however, discount historical research that may have been carried out under the banner of another subject. Geography for example was high on the list of those possessing and teaching computer skills.

<sup>5</sup> See H. J. Hanham: 'CLIO's Weapons', in: *Daedalus*, Spring 1971, pp. 509-19.

<sup>6</sup> *SSRC Survey Archive Data Catalogue: Guide to the Survey Archive's social science data holdings and allied sciences*, University of Essex, Colchester, (no date).

register data, insurance policies and marriage registers. A brief summary of these datasets follows:

- Records of Greenwood & Batley Ltd., machine tool producers, 1836-1900
- Aggregated demographic data from parish registers, 1537-1837
- Industrial and occupational data from 1911 census, Tayside.
- Marriage register data, nine Oxford parishes, 1837-1970.
- Marriage register data for London Kentish parishes, 1851-1853, 1874-1875.
- Aggregated demographic data, nine parishes in Tendring, Essex, 1538-1838.
- Migration in Northern Highlands, 1851-1891.
- Social structure in early 20th century Belfast. (census data)
- 1851 census, Norwich Wards, 1% sample.
- 1851 census of Seaham Harbour, County Durham.
- Censuses of 1851-1871, Sheerness Naval Dockyard
- 1861 census of Southwark and Christchurch, London.
- Index to 18th century fire insurance policy registers.
- 1851 census, Co. Antrim.
- Slave Ship records, 1817-1843.
- Slave Ship trade, 1791-1799 (from House of Lords survey).
- 1851 census, 2% national sample.
- Suffolk census data, 1851-1871.
- Ship tax for Harwich, Essex.
- Scottish Poor Law Commission statistics, 1844.
- British investors in the 16th and 17th centuries.
- Sample of insurance policies, 1750-1850.

Anybody can deposit data with the archive, indeed one family history society has already done so<sup>7</sup>. Probably more important is the fact that the ESRC requires that any data generated as a result of their funding should be deposited with the archive. However, unfortunately the situation is not as good as this requirement suggests. Those depositing data with the Essex Data Survey Archive, funded by the ESRC or not, form a clear minority of computer-using historians. The computerised datasets from many research projects, particularly post-graduate thesis work, have no doubt been deleted or lie disused in various departments throughout the country. It seems that

---

<sup>7</sup> The West Surrey Family History Society have deposited 1861 census data for Southwark, (Christchurch and St. Saviours).

most people, once having spent so much time and effort on collecting and making machine-readable a series of historical documents, wish to either safeguard the information from other people or gain something in return for their labours. Even for manual work, few of the many census indexes made by family historians have been deposited at either the Society of Genealogists or local record offices<sup>8</sup>. Ironically although the photocopying of documentary sources can raise considerable problems due to copyright, historians can make machine-readable or hand transcripts of them without any liability. Perhaps archivists should insist that copies of any transcriptions must be placed with the appropriate record office and/or other relevant data depository<sup>9</sup>.

It is easy to simply say 'send all machine-readable transcripts and data files to a data archive', yet the practicabilities of the situation are not quite so straightforward. The Essex Data Archive is already under-staffed and under-financed and may well not entirely welcome a flood of computerised transcripts. Even if the archive did have the facilities to handle such an influx the question, 'would people actually use these data files?' must be asked. There certainly seems to be a reluctance for academics to re-work 'second-hand' data and the degree to which local and family historians would want to use computerised data rather than selected print-outs of the information they require is questionable. What good is a library if no-one wants to read or borrow the books? This issue is of particular importance if the chief criterion for the funding of such an archive is the quantity of users rather than the quantity of depositors and quality of data. The question of data-quality leads on to what is probably the most significant problem area: how usable are the data files? The question of utility is very much interwoven with the whole issue of standards which has dominated this discussion<sup>10</sup>. However, it is not intended to rehearse the issues surrounding the standards debate in this article.

Historians in further education, as has been noted already, are no great users of computers, they probably understand the amorphous boxes even less. Worse still, the advice that many historians are offered shows little respect for the integrity of the historical source material. It is often assumed that a software package such as the much used Statistical Package for the Social Sciences (SPSS) will meet all of the historians computer requirements.

---

<sup>8</sup> J. Gibson and C. Chapman (eds.): *Census Indexes and Indexing*, Federation of Family History Societies, 1981.

<sup>9</sup> B. Collins: *The computer as a research tool*, in: *Journal of the Society of Archivists*, 7 (1), 1982, pp. 6-12.

<sup>10</sup> See especially *Computers in Genealogy*, vol. 1, no. 1 and vol. 1, no. 7.

In terms of the data file manipulation and statistical calculations available from SPSS this assumption may well be true, particularly in the case of the recently improved version<sup>11</sup>. However, the problem is that the traditional way of preparing information for this type of packaged analysis is to divide the data up into eighty character length records consisting of a number of numerically coded variables or fields<sup>12</sup>. Thus the inherent logic of the data is sacrificed to an obsolete computer technology. It is in this sense that the situation of historical computer-based research is possibly worse now than it was ten years ago. For example, of the few computer users a decade ago, several had acquired various programming and data-manipulation skills<sup>13</sup>, whereas today many historian computer users approach their research with the view that if it cannot be done within SPSS then it cannot be done at all. It is clear that such an approach is unsatisfactory.

Attempts have been, and are being made to remedy the situation and a framework for data analysis is being constructed. A clear distinction has to be made between data input and data collection. Data should be collected replicating the layout of the source as closely as possible, with logical records or units of variable length and secondary fields or data elements also of variable length<sup>14</sup>. The stage of data processing in which truncation, and coding occurs should be separated from data collection, with all processing being undertaken inside rather than outside the computer<sup>15</sup>. For documents of a standard format such as census returns, tithe awards and post-1812 parish registers, it is relatively easy to design a format which will represent the logical units of the original document in machine-readable form. A format has already been suggested elsewhere for the first of these documents<sup>16</sup>.

---

<sup>11</sup> N. H. Nie *et al.*: *SPSS: Statistical Package for the Social Sciences*, New York, 1975. Anon.: *SPSS-X User's Guide*, Chicago, Ill.: SPSS Inc., 1983.

<sup>12</sup> The main text book on the subject, although now very much out of date, recommends the use of numeric coding and restricting the record length to 80 columns. See E. Shorter: *The Historian and the Computer*, New Jersey, 1971.

<sup>13</sup> R. Schofield: *English Historians and the Computer*, in: *Historical Methods Newsletter*, 7, 1974, pp. 111-114.

<sup>14</sup> R. Schofield and R. Davies: *Towards a flexible Data Input and Record Management System*, in: *Historical Methods Newsletter*, 7, 1974, pp. 115-124.

<sup>15</sup> R. Floud: *An Introduction to Quantitative Methods for Historians*, (2nd ed.), London, 1979. See especially pp. 202-210.

<sup>16</sup> K. Schurer: *Methodology: Recording Data from the Original Sources*, in: *Historical Social Sciences Newsletter*, 2, 1984, pp. 8-11 and K. Schurer: *Census Enumerators' Returns and the Computer*, in: *Local Historian*, forthcoming.

### Machine-readable burial register (Bradfield, Essex 1814)

50/BRADFIELD/ESSEX/ONE THOUSAND EIGHT HUNDRED AND FOURTEEN  
60/17/JOHN, SHIPLEY/BRADFIELD/APRIL 17TH/28/HY THOMPSON VICAR  
60/18/-, BLYTH/MANNINGTREE/APRIL 24TH/4 YE/HY THOMPSON  
60/19/JOHN, SEABORN/BRADFIELD/APRIL 30TH/63/HY THOMPSON  
60/20/S-, FOX/BRADFIELD/JULY 1ST/76/HY THOMPSON  
60/21/MARY, TURNER/BRADFIELD/JULY 6TH/69/HY THOMPSON  
60/22/MARY, MORGAN/BRADFIELD/AUGUST 24TH/INFT/HY THOMPSON  
60/23/S-, GOYMER/BRADFIELD/SEPTR 11TH/39/HY THOMPSON

#### Example 1

while the other sources may be collected according to the formats illustrated in the following examples. A printed burial register may be recorded in a simple form shown in example 1. Lines tagged 50 indicate the start of a new page and lines tagged 60 relate to the individual row entries. The columns across this row are indicated by slashes (/), and logical subdivisions within these units such as between prenames and surnames are shown by a comma.

A similar framework can be applied to a tithe award, as in example 2.

The lines tagged 59 and 69 indicate the column headings applicable to the fields in the lines tagged 50 and 60 respectively. Each parcel of land starts with a line tagged 50 and ends with a subsequent line 50 or a line tagged 70, which totals the quantity and amounts payable for the particular holding. Separate land divisions within a single holding (i.e. fields) are detailed on lines tagged 60. Columns with no information are represented by a single dash (-) unless they are trailing, in which case they are omitted.

However, the interests of historians are not confined to documents of a standard form and logical consistency; wills, deeds and even early parish registers all have little or no coherent logical structure. Faced with this predicament computer using historians have often adapted the data to fit their programmes by juxtaposition, abbreviation and retention of only the information that interests them; thus rendering the dataset near worthless for the purposes of archival use. A much wiser approach to this problem would be to collect the data in a free-format fashion replicating the source, change the programmes and let the computer carry out any adaptation, retaining at all time a machine-readable copy of the source which can always be referred to or re-worked as necessary. Unfortunately, computers cannot readily understand text typed into them verbatim, in order to produce the sort of analyses historians require the computer has to be given a sense of location. This

Machine-readable tithe award (Steeple, Essex 1839)\*

59/LANDOWNERS/OCCUPIERS

69/NO/NAME/STATE/QUANTITY/PAYABLE TO VICAR/PAYABLE TO J K +  
HUNT & ELIZ HUNT & THOMAS HUNT

50/EXORS OF THE LATE ISAAC,BROWN/JOHN,RADLEY

60/12/KINGS 6A/ARABLE/6,2,22/-/1,15,10

60/13/ROADFIELD/DO/6,2,9/-/1,17,-

60/14/6 1"/"2A/DO/6,3,14/-/1,18,6

60/15/4 A/DO/3,3,34/-/1,3,1

60/16/3 A/DO/3,3,37/-/1,3,5

60/17/4 A/DO/4,-,28/-/1,1,6

60/18/BARNFIELD/DO/3,3,5/-/1,1,6

60/19/GARDENFIELD/DO/3,1,16/-/1,18,3

60/20/HOMESTEAD/DO/,2,3

60/21/2 A/GRASS/2,3,25/-/-,3,-

60/22/6 A/ARABLE/6,1,4/-/1,13,2

60/23/5A MARSH/DO/5,1,37/-/1,6,9

60/24/3A MARSH/GRASS/4,-,38/-/-,5,-

70/58,2,32/-/14,7,-

50/WILLIAM, BLAKE/GEORGE, PATTISON

60/202/LOST FIELD/ARABLE/13,2,15/1,6,-/3,-,3

50/, SOCIETY OF QUAKERS/, SOCIETY OF QUAKERS

60/218/BURIAL GROUND/GRASS/, ,31

\* Note that slash in 6 1/2 A (No 14) has been surrounded by quotation marks to distinguish it from a slash separating data elements. Also lines longer than 80 characters are continued on a subsequent line and the continuation indicated by a plus sign (+) on the continued line and by indenting the continuation line two spaces. Dashes (-) in the amounts payable and sizes of fields are permissible since these cannot be confused with dashes indicating missing data elements as these are always between two slashes.

Example 2

may be achieved by inserting a series of unique flags or pointers into the text indicating the presence of words or phrases which may be considered by the researcher to form logical data elements or variables<sup>17</sup>. For example, an early parish register may be represented in machine-readable form as in example 3; baptisms are recorded on lines tagged 61, burials on lines tagged 62, marriages on lines tagged 63 and prenames, surnames and dates are all flagged, flags starting and ending with an asterisk.

Alternatively, according to the structure of a particular document, the researcher may wish to interleave lines of free-field and fixed-field formats. For example, in the baptism register of St. James, Clerkenwell the incumbent generally made the entries in the format as follows: month; date; prename; " 's' or 'd' of "; father's prename; parents surname; "&"; mother's prename; "his wife; born"; date. Therefore, lines entered in this standard fashion may be tagged to indicate that they are in this fixed format, whilst lines breaking this convention can be tagged and flagged accordingly as in example 4.

However, although in this example slashes separating logical elements are substituted by spaces, such practice is potentially dangerous since if the recording of a middle name or some other detail went unnoticed, the extra space would throw the computer into disarray.

Unfortunately flagging of data elements is not a straightforward task. If all that is required for retrieval purposes is the flagging of names and occupations then the necessary flags can be quite simple. However, if the researcher wishes to link data elements and retain the context in which the data elements occur, then the process of flagging can become extremely complicated. This latter point is illustrated by work carried out by Alan Macfarlane and colleagues in a project aimed at making machine-readable every document over the period 1550 to 1750 relating to the Essex village of Earls Colne<sup>18</sup>. Wishing to retain the grammatical syntax of the text, the documents were broken down into a series of entities based round a subject matter. These were then flagged, bracketed and linked to other entities by nesting and numbering of brackets<sup>19</sup>. As example 5 shows, for a parish register the scheme

---

<sup>17</sup> See M. Overton: Computer analysis of an inconsistent data source: the case of probate inventories, in: *Journal of Historical Geography*, 3 (4), 1977, pp. 317 - 326 and G. A. Dobbert: An On-line System for Processing Loosely Structured Records, in: *Historical Methods*, 15 (1), 1982, pp. 16-22

<sup>18</sup> The research project is based at the Department of Social Anthropology, University of Cambridge.

<sup>19</sup> C. J. Jardine and A. D. J. Macfarlane: Computer Input of Historical Records for Multi-Source Record Linkage, in: M. W. Flinn (ed.): *Proceedings of the 7th*

Machine-readable Parish Register (Bradfield, Essex 1738)\*

50/1738

61/BAP: \*P WILLIAM\*, THE SON OF \*P SAMUEL\* AND \*P MARTHA\* +  
\*S CARRINGTON\*, WAS BAPTIZED THE \*D 4TH OF SEPTEMBER\* . +  
1738

61/BAP: \*P SARAH\* AND \*P MARY\*, DAUGHTERS OF \*P THOMAS\* AND +  
\*P BRIDGET\* \*S BRASSTREE\* WERE BAPTIZED THE \*D 8TH OF +  
OCTOBER\* .

62/BUR: \*P SARAH\* & \*P MARY\* \*S BRASSTREE\* WERE BURIED THE +  
\*D 16TH OF OCTOBER\* THE AFFIDAVIT FOR BURYING IN WOOLLEN +  
REGISTERED

61/BAP: \*P ELIZABETH\* , THE DAUGHTER OF \*P JOHN\* AND \*P +  
SUSANNA\* \*S ROWLAND\* WAS BAPTIZED THE \*D 22D OF OCTOBER\* . +  
1738

61/BAP: \*P JOSHUA\* , THE SON OF \*P JOHN\* AND \*P ANNE\* \*S NUN\* +  
, WAS BAPTIZED THE \*D 5TH OF NOVEMBER\* . 1738

62/BUR: \*P SARAH\* \*S YELL\* WAS BURIED THE \*D 14TH OF NOVEMBER\* +  
. A.R.

62/BUR: \*P SUSANNA\* \*S PLUMMER\* WAS BURIED THE \*D 15TH OF +  
NOVEMBER\* . A.R.

63/MAR: \*P THOMAS\* \*S GOSS\* AND \*P ELIZABETH\* \*S KING\* WERE +  
MARRIED THE \*D 16TH

OF NOVEMBER\* , THEIR BANNS BEING FIRST THRICE DULY PUBLISHED

62/BUR: \*P ALICE\* \*S KING\* WAS BURIED THE \*D 17TH OF DECEMBER\* +  
. A.REGIST. BEING BROUGHT WITHIN EIGHT DAYS, THE TIME +  
PRESCRIBED BY THE ACT OF PARLIAMENT, MADE FOR BURYING IN +  
WOOLLEN.

\* Note that in this example an asterisk is used to both end and start a flag.  
Lines longer than 80 characters are continued on a subsequent line and the conti-  
nuation indicated by a plus sign (+) on the continued line and by indenting the  
continuation line two spaces.

\*P = Prenom

\*S = Surname

\*D = Date

Example 3



Mixed free and fixed formats (St James, Clerkenwell 1698)\*

60/JUNE 16 MARY D. OF WILL. COOKE & MARY HIS WIFE; BORN 24 MAY  
 60/JUNE 19 MARY D. OF JAMES NEWTON & ELIANOR HIS WIFE; BORN 6  
 60/JUNE 24 ELIZ. D. OF JOHN STORER & ELIZ. HIS WIFE; BORN 24  
 60/JUNE 24 RICH. S. OR ROBT SAVAGE & FRAN. HIS WIFE; BORN 20  
 60/JUNE 26 WILL. S. OF JOHN MUNDAY & ANN HIS WIFE; BORN 13  
 60/JUNE 26 WILL. S. OF JOHN JENKINSON & ELIZA. HIS WIFE; +  
 BORN 13  
 60/JUNE 26 SARAH D. OF RICH. SNOW & ANN HIS WIFE; BORN 18  
 70/\*D JUNE 27\* \*CN CATH.\* D. OF \*FN WILL.\* & \*MN CATHR\*  
 \*S SMITH\* , BUT I BELEEVE THIS IS ILEGITIMATE  
 60/JUNE 28 WILL. S. OF THO. ORETON & ANN HIS WIFE; BORN 10  
 60/JUNE 30 WILL. S. OF WILL. PAGE & SARAH HIS WIFE; BORN 29  
 60/JULY 3 JOSEPH S. OF JOHN BARRETT & SARAH HIS WIFE; BORN +  
 20 JUNE  
 60/JULY 3 THOMAS S. OF ISAAC SUFFOLKE & ELIZ. HIS WIFE; BORN +  
 29 JUNE  
 60/JULY 3 WILL. S. OF WILL. DRAKE & CATHR HIS WIFE; BORN +  
 23 JUNE  
 60/JULY 5 SARAH D. OF DANIEL KENEDAY & JONE HIS WIFE; BORN 5  
 70/\*D JULY 15\* \*CN THO.\* S. OF MR \*FN DEUEL\* \*S PEAD\* , +  
 GENT. , & \*MN SARAH\* HIS WIFE; \*BD BORN 4\*  
 60/JULY 17 SAMLL S. OF JOSHUA ATKINSON & FRAN. HIS WIFE;  
 BORN 14  
 60/JULY 17 DOROTHEA D. OF HENRY DOWNER & ELIZ. HIS WIFE;  
 BORN 5  
 60/JULY 21 JOHN S. OF ROBT NOBLE & SARAH HIS WIFE; BORN 21  
 60/JULY 22 ELIZ. D. OF JOHN KINGSTON & ELIZ. HIS WIFE; BORN 13

\* Example taken from R. Hovenden (ed): A True Register of all the Christenings, Mariages and Burialles in the Parishe of St James, Clarkenwell, From the Yeare of our Lorde God 1551, Vol 1 Christenings, 1551 to 1700, The Harleian Society, (Registers, (Volume IX)), London, 1884, p.376 .

\*D = Date (of event)

\*CN = Christian Name

\*FN = Father's Name

\*MN = Mother's Name

\*S = Surname

\*BD = Birth Date

Example 4

## Example of linked flagging (from Earls Colne, Essex)\*

### A: Parish Register

Source text:

JOHN THE SON OF HENRY ABBOTT WAS BAPTISED 5TH MAY 1607

Logical structure:

[A person] [who has a name] [and who is involved in a kinship relation] [with another person] [who has a name]. [The first person is involved in an event] [on a date].

Input data:

(P (N JOHN ) (K THE SON OF (P (N HENRY ABBOTT ) ) )  
(E WAS BAPTISED (D 5TH MAY 1607 ) ) )

### B: Will

Source text:

HENRY ABBOTT AND JONE HIS WIFE DO CLAIM FOR HOLD A TENEMENT  
IN CHURCH STREET

Input data:

(P \*1 (N HENRY ABBOTT ) ) AND (P (N JONE ) (K (1 HIS ) WIFE )  
(H DO CLAIM FOR HOLD (L A TENEMENT IN CHURCH STREET ) )

\* P - Person

N - Name

K - Kinship relation

E - Event

D - Date

H - Landholding

L - Description of land

Example taken from *C.J.Jardine and A.D.J.Macfarlane: Computer input of Historical Records for Multi-Source Record Linkage*, In: *M.W.Flinn (ed): Proceedings of the 7th International Economic History Congress, Edinburgh 1978*, 2, Edinburgh, 1978, pp.71-78

### Example 5

is relatively easy to implement, however, a lengthy will may require dozens of nested and linked brackets which can only be added after the document has been read, understood, broken down into its composite entities, and the entities linked with each other.

Regardless of the simplicity or complexity of the flags and pointers used by a researcher it is important that a number of points are adhered to. The combination of characters used as flags obviously have to be unique and must not occur in the text of the document. Yet equally a flag should not be too long, its meaning should be implicit, perhaps mnemonic, and it must be portable. Therefore, it must not contain characters that cannot be reproduced by other computers, printers or typewriters. Also it must be remembered that flags are only used as pointers to words or phrases that the current researcher feels may be of interest. No set of predetermined pointers will be absolute, other researchers will wish to sub-divide singularly flagged data elements and re-define or re-classify various flagged data elements altogether. Therefore, if data are to be stored in a data archive it is of crucial importance that any flags used do not change the shape of the data and that the original text of the document is always retrievable. It is for these reasons that in the case of free-format documents it is probably best to initially type in the text with appropriate tags indicating the structure of the document, and then add any required flags at a secondary editing stage, thus retaining the flag-free version for reference and archive purposes. Such a policy would also satisfy the demands of the many non-computer orientated historians who just want a legible, understandable line-printer copy of the data. Additionally, if those persisting with manual research prepared their transcriptions in line with the conventions used in the first of these stages, then high quality transcriptions could be fed into a computer via an optical character reader, such as a Kurzweil Data Entry Machine (KDEM) and then flagged and formatted as required.

Many of the points that have been made look slightly to the future, to a situation where data can be referred to, deposited with, and taken out of an archive in much the same way as we currently use libraries. Nothing has been said about the host of technical problems surrounding the accomplishment of this situation, about the problems of incompatibility of storage media and storage formats. In the past changes in technology have adjusted the nature and intensity of many of these problems and will, no doubt, continue to do so in the future. For example, an interesting recent innovation has been the development of the multi-format floppy disc copying micro-computer, called the 'magic machine' for short, which, capable of handling over seventy different storage formats, has been installed at the Essex Data Archive to offer a disc transfer service to those who are unable to transfer information between micro-computers at their own institutions. However, regardless of the technical abilities, such a service would be dramatically devalued if the

data being transferred could not be historically interpreted as a result of data reduction or subscription in the form of various codes and classifications.

Still looking towards the future words of caution have to be sounded not only about historical sources being converted into an historical database, but also about the data that is being created now for the historian of the future. To what use will the historian put the machine-readable documents that our society is currently creating? Will he be able to use them at all? Already the discs on which the 1960 American census data are stored cannot be read because the disc drives are no longer made and all of the old ones have been scrapped. Similarly the magnetic tapes holding the British 1961 census have already decayed to the point that researchers wanting special tabulations from them have been refused for "technical reasons". What of more everyday archives; the accounts of small-businesses, the wage-books of companies and the records of schools and local government? Will the historian of the future be able to read such files on his computer, or will he have to use a specialized machine, or will all the files he wishes to consult have been written-over as they become out-of-date anyway? What will the scientific or literary historian make of films containing no drafts, just mathematical formulae and prose, neatly edited with all trace of initial, superseded ideas deleted? All of these questions demand an answer which needs to be considered very carefully.



## Computing the History of the Family: A Question of Standards

### The present state of research

Over many years the Cambridge Group has built up a valuable collection of censuses and listings and, from these, produced sets of tabular analyses (in manuscript) displaying the family and household structure of various communities. The listing file, as it has come to be called<sup>1</sup>, comprises a collection of most known enumerations of English communities prior to 1801 which divided the population into name blocks which can be identified as residential units of some sort although the precise definition of that unit often remains unclear<sup>2</sup>. Holdings of censuses dating from after 1801 and listings of overseas populations are selective, reflecting the individual interests of members of the Cambridge Group. In all, approximately 500 censuses of English communities have been included in the collection<sup>3</sup>; not very many when one considers that pre-industrial England was divided into approximately ten thousand ancient parishes.

Arising from the fact that the collection has taken a number of years to assemble, it is inevitable that the analyses referred to above should vary in accuracy, in the conventions followed (in regard, for example, to the iden-

---

<sup>1</sup> P. Laslett: The study of social structure from listings of inhabitants, in: E.A. Wrigley (ed.): *An Introduction to English Historical Demography*, 1966, 160-78. The procedures for analysing these documents have subsequently been substantially revised cf. P. Laslett: Mean household size over three centuries, in: P. Laslett and R. Wall (eds): *Household and Family in Past Time*, Cambridge, 1972, 127. In addition, much stricter criteria are now employed in the selection of documents, particularly when the research focus is on the composition of the household in terms of the relationships of the members to the household head. See R. Wall: The household, demographic and economic change in England, 1650-1970, in: R. Wall, J. Robin and P. Laslett (eds): *Family Forms in Historic Europe*, Cambridge, 1983, 496 n9. The unpublished listings analyses are available for consultation in the library of the ESRC Cambridge Group for the History of Population.

<sup>2</sup> These issues are explored in greater detail in the introduction to Wall, Robin and Laslett, 1983, 6-13.

<sup>3</sup> A full list of the places has appeared in successive issues of the journal *Local Population Studies* beginning in 1968. From *Local Population Studies* 24 (Spring 1980) the list has been repeated and expanded to include details on the range of information on individuals and households to be found in each document.

tification of households)<sup>4</sup> and in the degrees to which the listing has been fully exploited as fresh tables have been added to the series and old ones redrafted. For some years we have been working on the construction of a set of model tables, similar to those recommended by the United Nations for the presentation of current statistics on the family and household but suitably adapted to reflect the range of persons to be found in households in past societies by comparison with more contemporary societies. The tables, some 85 in all, are divided into various sections according to the particular aspect of the population to be investigated: children, the elderly, servants, relatives, boarders and lodgers, employment and class structure, and migration. Two preliminary sections deal with broader concerns such as the distribution of the population by age, sex and marital status and the composition of the household in terms of both kin structure and the balance between producers and consumers. The objectives are threefold: to move discussions of household forms from their obsessive preoccupation with the kin structure of households<sup>5</sup>, to recommend standard definitions of critical terms, for example 'family', 'household', 'child', 'relative' etc., and to achieve a minimum level of detail in the relevant tabulations. To the latter end the Cambridge Group, for the cost of computer print-out plus postage<sup>6</sup>, will make available

---

<sup>4</sup> One of the principal difficulties is whether to count lodgers and their families as constituting separate households. The definition of the household favoured by the Cambridge Group is to see lodgers as constituting appendages to the household (cf. the introduction by *P. Laslett* to: *Household and Family in Past Time*, 1972, 34-9) but this in turn gives rise to a further problem in that strictly speaking the situation in which a lodging group contains children but the main part of the household does not, constitutes a household without children. However, there is reason to think that this decision was not applied consistently as listings were processed by a series of researchers at the Cambridge Group.

<sup>5</sup> Cf. the seemingly never-ending debate on the dominance or otherwise of the nuclear and stem family. The cause of the stem family was championed initially by *Berkner*. See: *The stem family and the development cycle of the peasant household*, in: *American Historical Review*, 77/2, 398-418. *Laslett's* reply came eventually in a contribution to *K.W. Wachter, E.A. Hammel and P. Laslett* (eds): *Statistical Studies of Historical Social Structure*, London: Academic Press, 89-111, but the debate rumbles on particularly in the pages of the *Journal of Family History*. For a recent contribution see *Harvey Smith*: *Family and Class: the household economy of Languedoc wine growers, 1830-1870*, in: *Journal of Family History*, 9/1 (Spring 1984), 65 n.1.

<sup>6</sup> A list of the table titles appears in the Appendix to *P. Laslett*: *The family and*

to interested researchers any or all of the 85 tables in a form suitable for the insertion of their own results.

The tables were initially tested on three small populations: English, Corsican and Russian, selected because it was felt each would be dominated by quite distinct types of family; such indeed proved to be the case. However, the full potential of the model tables lies in the exploration of the inter-relationships of household composition with inheritance custom and practice and particular types of economy and with this objective our research will in the future focus on the better documented communities of mid-nineteenth century Britain and on selected overseas communities. The social structure of mid-nineteenth century Britain is surprisingly under-researched despite censuses of high quality in which individuals are identified by name, age, sex, marital status, relationship to household head, occupation and county and parish of birth. One page from the census enumerator's return for the Dorset parish of Corfe Castle in 1851 is reproduced in Figure 1. In making his return the enumerator failed to keep to the guidelines issued by the census authorities in London (cf below Figure 2 for a transcription and commentary on the errors) but the general format is clear and uniform across the country for a given census year. Censuses survive for virtually every place in the country beginning in 1851 and are currently open for public inspection up to 1881<sup>7</sup>. Despite this, informative studies number only a handful<sup>8</sup> and are al-

---

household as work group and kin group: areas of traditional Europe compared, in Wall, Robin and Laslett: *Family Forms in Historic Europe*, 1983, 560-3. Further details are available from Kevin Schurer at the Cambridge Group, 27 Trumpington Street, Cambridge CB2 1QA.

<sup>7</sup> Censuses of England and Wales have been held decennially since 1801 with the exception of 1941. Returns of the first four censuses survive in some numbers of in local archives but vary in quality. An almost complete set of returns exists for 1841 but no attempt was made to record marital status and relationship to the head of the household while the ages of adults were rounded down to the nearest five year point (i.e. a man aged 62 with a wife aged 59 would have had their ages recorded as 60 and 55 respectively).

<sup>8</sup> Dennis Mills and Carole Pearce have produced a usefull bibliography which details the subjects covered in studies based on the unpublished enumeration returns of the mid-19th century censuses. This appeared first as: *Census Enumerators' Books* (1982), Faculty of Social Sciences, The Open University, but has since been extended and updated as a computerised bibliography in a joint project with the Cambridge Group. Further information is available from the Group's Data Editor, 27 Trumpington Street, Cambridge CB2 1QA.



most exclusively concerned with 1851. Research on the censuses of 1861-1881 is totally uncoordinated and is often of dubious value. As for the mapping of European familial 'regions', designed to establish the principle characteristics of the East European, North-West European and the Mediterranean 'family', the approach has been to depend on the analysis of a few arbitrarily selected communities, widely scattered in terms of both historical time and geographical location. These are often then formulated by Peter Laslett in the late 1960s<sup>9</sup>.

The existence of a set of model tables, however, will not of itself result in a rapid accretion of knowledge so long as the tables have to be completed by hand since the task can last some months even for a moderately sized parish. An obvious solution is to process the census by computer. In the 1970s, Peter Laslett was recommending data exchange on household by means of the ideographic system<sup>10</sup>: now data are available on magnetic tape but this would be no advance if the data cannot be satisfactorily interpreted. The successful interchange of computerised data requires a standard data collection format, procedures for processing and tabulating data on households and families, and, of course, the set of model tables already referred to so that however different the researchers objectives, the results are strictly comparable. These three requirements can now be met.

### Standard Input Format

The proposed standard format for data collection is depicted in Figure 2 using as an example the same page from an enumeration return that was represented in Figure 1. The prime objective is to follow as closely as possible both the content and layout of the original document, avoiding both pre-coding and abbreviation. In other words, one is producing a transcript in machine-readable form and not a series of codes substituting for a transcript. Not only will this speed the input process but it eliminates the inevitable errors when coding accompanies or precedes input. Naturally, some input errors will occur, but the input of text rather than codes makes checking and rectification of fatal errors that much easier. The present example relates to

---

<sup>9</sup> Cf. introduction by Laslett to: *Laslett and Wall: Household and Family in Past Time*, 1972, 74-85 and *Laslett: Characteristics of the western family considered over time*, in *Laslett: Family Life and Illicit Love in Earlier Generations*, Cambridge, 1977, 22, 26.

<sup>10</sup> *Laslett: Introduction*, in: *Laslett and Wall: Household and Family in Past Time*, 1972, 41-4.

census but any document containing standard information on individuals or families, such as a manor court roll or parish register, could in principle be handled in a similar way by applying the same logic of input<sup>11</sup>. Additions are made to the text to indicate the structure of the document. These involve the insertion of oblique slashes as field delimiters, the insertion of commas to denote subdivisions within fields as, for example, between prename and surname, and a numeric tag to indicate the nature of the information contained within each line. In the present example (see Figure 2) all lines prefixed by the tag 50 relate to a household and all lines prefixed by tag 60 to an individual. Also a plus sign is added to carry over a line of text beyond the maximum 80 characters. All the information available on that individual is recorded while the household line contains the schedule number and the address. As the household line denotes the start of a new household, it must always be present even if there are no schedule numbers and no addresses as is usually the case with pre-industrial listings.

Other types of line have also been defined but occur less frequently. Lines beginning with 70 indicate the end of a house or houseful as we have come to term it<sup>12</sup>. In the census of 1851 this is indicated by a line drawn across the whole page of the enumeration book indicating that the residents of a new house are being listed<sup>13</sup>. Comments made by the researcher can be included in lines commencing with 81-89. These lines may also be used to record occasional or semi-structured information such as page headings<sup>14</sup>. Comments are always confined within double parentheses to set them apart

---

<sup>11</sup> See K. Schurer: Historical Databases and the Researcher, in: Computers in Genealogy (forthcoming).

<sup>12</sup> Cf. Laslett: Introduction in: Laslett and Wall: Household and Family in Past Time, 1972, 35-6.

<sup>13</sup> Procedures for identifying households in the 1851 enumeration returns are set out by Michael Anderson in Standard Tabulation Procedures for the Census Enumerators' Books 1851-1891, In E.A. Wrigley (ed.): Nineteenth Century Society, Cambridge, 1972, 134-45. Peter Tillott in the same volume points to some of the pitfalls that await the unwary researcher, see: Sources of inaccuracy in the 1851 and 1861 censuses: 82-133.

<sup>14</sup> In Figure 2 the totals of houses, and of males and females as given by the enumerator at the bottom of the page (cf. Figure 1) appear in line 82. Line 89 is used on five occasions to record both errors in the enumeration and corrections by the enumerator or others. This is required since such cases may cause difficulty in the drawing of boundaries between households or the interpretation of characteristics of a particular individual.

from the text of the document. In addition, a series of statements is included at the commencement of each file, labelling and defining the file in terms of its number, time of creation, character (whether for example it is a census or some other document), detailed description of the community to be investigated and the archive reference number or numbers of the document.

Fields are of variable length but recorded in fixed format, that is the fields of information must follow each other in a strict order. In the present case (Figure 2), the order is as follows: name (subdivided into prenames and surname), relationship to household head, marital condition, male age, female age<sup>15</sup>, occupation, and birthplace (subdivided into community, county and country hierarchies)<sup>16</sup>. Censuses containing less or more detail on individuals or households or simply presenting the same information in a different order each require a different format. However, so long as the processor or the data is appraised of the content of each field, this should cause no problem. Furthermore, the choice of a limited and an internationally recommended character set means that texts submitted in this form can be read into different makes of computers<sup>17</sup>. Interchange between computers is also facilitated by not allowing physical line length to exceed 80 characters.

### The input process

The various stages of data checking, coding and correction are summarised in Figure 3, based directly on our experience at the Cambridge Group in the processing of census data from Britain, Austria and Turkey<sup>18</sup> using a

---

<sup>15</sup> An individual's sex was not separately recorded. Instead the age was entered as appropriate into a column for males or for females.

<sup>16</sup> The final column of the enumeration schedule (cf. Figure 1) for the recording of a person who was blind, deaf or dumb is almost invariably blank and is not shown in the transcription.

<sup>17</sup> British Standards Institution, Draft for Development: the structure and representation of data for interchange at the application level (DIAL), Part 1, recommendation for syntax and basic principles, DD75, Part 1 Appendix C (1981).

<sup>18</sup> Machine-readable data for ten Welsh communities in pre-processed form corresponding to the public data format, was supplied by Michael Anderson from his nationwide sample of enumeration districts in 1851. The project is described in *M. Anderson et al., Preparation and analysis of a machine-readable National Sample from the Enumerators' Books of the 1851 Census of Great Britain. A final report to the Social Science Research Council, HR2066 (1980).* The public data format and the advantages of this method of data transmission over the data as collected

computer package called the Statistical Analysis System (abbreviated henceforth to SAS). After punching the census a copy of the raw data is sent to store for security and future reference. In Cambridge the store medium is tape (Figure 3, 1-3). Alternatively, data already on magnetic tape may be acquired directly from other researchers. In these circumstances, it is possible to proceed directly to stages 4 and 5, the creation of a SAS dataset and an inventory of all the items occurring in each field. The inventory is compiled in reference to a coding dictionary of relationships, occupations and marital conditions, abstracted from censuses processed earlier or created from partly or fully coded variables assigned by the data collector. Any items not located in the coding dictionary are output in order that the researcher may assign an appropriate code and update the dictionary. On resubmission all codes are by definition located and can be copied from the dictionary onto the main file and, for safety, copied to tape (Figure 3, step 7-8). The encoding and checking (step 6) is a most important element in the input process in that it provides an opportunity to look for solutions for the incompatibilities between codes. The latter for example would arise if a person had been assigned a code indicating that she was a wife but was reported as being unmarried.

The same programmes also output any entries which seem likely to cause problems of one sort or another in the production of the tables. Some of these issues arise because there may be reason to suspect the accuracy of the date, an extensive age gap between husband and wife, or between mother and child for example. Other difficulties stem from the fact that some of the tables require information on relationship between certain members of the household as well as between each member and the household head, and this is not always self-evident from the given relationship. To take a case in point, whether a grandchild of the head lives with his parents as well as his grandparents may on occasion be inferred from the order in which the individuals are listed in the household while in other cases the situation is uncertain as long as a representative of the intervening generation is present of an age to be the parent and there is no conflict in surname. Others may not agree on the interpretation of such difficult cases and for this reason it is important to preserve as much flexibility as possible. This has been achieved

---

(the data collection format) is described by *John Welford*: The establishment of portable interchange formats for genealogical data – can we hope to reach an acceptable standard?, in: *Computers in Genealogy*, 1/7 (March 1984), 178-87. *Dr. Manfred Thaller* (Göttingen) and *Dr. Ehmer* (Vienna) supplied the data on a suburb of Vienna in 1880 and *Dr. Cem Behar* and *Dr. Alan Duben* provided the data on 19th century Istanbul.

by preserving the old cases along with the new and these can be activated to produce varying estimates of the number of extended and multiple households while the raw data always remain available for reference. Other checks allow the researcher to identify all cases of married people whose spouse was absent on census night and servants and lodgers who shared a surname with the head of the household and ought to be considered as possible relatives instead of, or in addition to, their status as servants or lodgers.

When all the checking and correcting of the files has been completed, and this usually involves a further run through the correction phase to ensure that all difficulties have been considered and no fresh errors introduced (steps 6 and 9), the revised SAS files are copied to store (Figure 3, step 10) to constitute the working data library. This library will be accessed by a further series of programmes to generate the Model Tables.

### The production of the tables

The SAS working data library is a partitioned file and consists in fact of four members. SPBASE (cf. Figure 4) contains all the information on individuals while HOUSE contains all the information on the residential location for the houseful and household. Both members are retained primarily for reference purpose. The member SP is developed from SPBASE and contains the variables and their coded values required for the production of the tables whenever the reference point is the individual rather than the household. The member HH fulfills a similar purpose for households. It contains similar information to the SP file but is organised at the household level rather than at that of the individual, i.e. it holds all information on people in the household, and can be used to count the number of children, relatives and conjugal family units per household. These four data library members form the core of a computerised census analysis system (CAMTAB) which has been developed at the Cambridge Group.

An additional element to the CAMTAB system is a procedure which assigns the Laslett-Hammel classification of household structure to each household. This six-fold classification into *solitaries*, *no family*, *simple*, *extended*, *multiple*, and *indeterminate households*, first outlined in 1972, has been elaborated but remains essentially unchanged in the present formulation<sup>19</sup>. The

---

<sup>19</sup> See below Figure 5 and cf. Laslett: Introduction, In: Laslett and Wall: Household and Family in Past Time, 1972, 31. The principal differences concern the more detailed specification in the revised table of the no-family and extended family households.

basic unit is the simple family household of no more than one married couple with or without children, or one lone parent-child group. Extended households are formed by the presence in addition of relatives drawn from outside the immediate nuclear family and the household is considered multiple if these relatives in turn form a secondary marital or parent-child group. Less complex households are represented by persons living alone (solitaries) and by groups of related and unrelated people co-residing, none of them married or being a member of a parent-child group. Each household type is then subdivided to show for example whether the solitaries are widowed or unmarried, whether the simple family households consist of married couples with or without offspring or widows or widowers with offspring and whether extended family households are extended upwards, downwards or laterally from the core marital or parent-child group. All these possibilities are spelt out in Table B3 of the model series (see below Figure 5).

The programme classifies households by type first eliminating lodgers, servants and visitors from consideration and then proceeding to identify in turn households of solitaries, no family households and simple family households using such information as the number of persons present, their marital status and their relationship to each other and to the head of the household. Much use has been made by researchers of the Laslett-Hammel classification scheme, and the trained eye can almost instantaneously distinguish an extended from a no family or simple household. However, specifying the logic of these decisions has proved considerably more arduous because all the possibilities must be considered. Figure 6, therefore, is limited to an illustration of the procedures to be followed in identifying the first three household types: solitaries, no family households and simple family households. Together with the further sections which cover extended and multiple family households, these procedures constituted the core of the programme that successfully classified all the households in 10 Welsh communities enumerated in 1851 and nineteenth century Vienna and Istanbul (cf note 18). No doubt, it is possible, indeed likely, that some further modifications of the programme may be required if communities experiencing a yet more complex household type were to be analysed.

## Conclusion

Considerable attention has been devoted above to Table B3 because the classification of households in terms of their kin structure has in the past been seen as the arbiter of whether different populations share the same family system. However, further research soon established that there were many

other aspects of the household, for example the presence of servants, the size of the age gap between husband and wife and the changes in the number and type of persons present according to the age of the household head, that were equally deserving of attention<sup>20</sup>. It was to meet such points that the series of model tables was developed and they represent our view of what ought to be involved in a full interpretation of the structure of the households. Nevertheless, it is not to be expected that it will prove possible or even desirable to complete all the tables in every community that is investigated. Nor is it possible to circumvent all the problems of analysing families and households at a point in time when it is well known that their structure is subject to continual change. The process of following individuals, families or households from census to census requires quite a different methodology, and one that is not yet fully developed<sup>21</sup>. Even in terms of the snapshot picture of the household, the set of eighty-five tables may well be insufficiently comprehensive.

Only two claims in their support have been advanced here. The first is that they provide a wide range of standard information on the family and household which individual scholars can consult to place any particular community study in appropriate context. Secondly, the tables serve notice of our intention to end the excessive attention devoted to the kin structure of the household and the tendency to view it in isolation from other features of the society such as employment patterns. In the future, therefore, we will be looking not just at the kin structure of household but whether the more complex households were more or less likely than other types of household to contain servants or inmates. The role of the household as an economic unit will be stressed with a variety of analyses of the size, age and sex composition of its resident labour force. However, it is far from our wish that such lines of enquiry should totally replace the investigation of the kin structure of households. On the contrary, it is the precise relationship between this kin structure and the nature of the local labour market that will be the focus of attention.

---

<sup>20</sup> See especially Wall: The composition of households in a population of six men to ten women: south-east Bruges in 1814, In: Wall, Robin and Laslett: *Family Forms in Historic Europe*, 1983, 460-72.

<sup>21</sup> Cf. Wall: Introduction, In: Wall, Robin and Laslett: *Family Forms in Historic Europe*, 1983, 4, commenting on the different approaches of Reinhard Sieder and Michael Mitterauer and Luc Danhieu in their respective contributions to the same volume.





10/DATA COLLECTION FILE/C1000/K.SCHURER/3 JULY 1984  
 20/HOUSEHOLD CENSUS/1851  
 31/6073/STANDARD/RURAL  
 32/ENGLAND/DORSET/CORFE CASTLE  
 41/PRO BO 107 1856 130-162  
 42/ALL THAT PART OF THE PARISH OF CORFE CASTLE WHICH LIES TO THE SOUTH OF +  
 CHALLOW HILLS AND EAST OF THE STREAM RUNNING FROM BLASHENWEL TO WEST MILLS +  
 INCLUDING WILLIAM SMITHS AND ISAAC PHILLIPS' AT THE BRIDGE REV G.HUBBARD +  
 BREAKNECKS CHALLOW SANDY HILLS WOLLGARSTON AILWOOD WESTWOOD TAPERSHILL EAST +  
 TOWNSEND AND EAST STREET ''/' BOTH SIDES BOAR MILL ROUND THE MARKET PLACE +  
 AND WEST STREET ''/' BOTH (SIDES) TO WEST TOWNSEND  
 .  
 .  
 81/CORFE CASTLE/DIOCESE OF SALISBURY/((?))/CORFE CASTLE/((2))  
 50/51/EAST STREET  
 60/HENRY, DAY/HEAD/MAR/36/-/JOURNEYMAN CARPENTER/DORSET, CORFE CASTLE  
 60/ELIZABETH, DO/WIFE/MAR/-/31/-/DO, DO  
 60/GEORGE, DO/SON/-/14/-/SCHOLAR/DO, DO  
 60/ALBERT, DO/SON/-/8/-/DO/DO, DO  
 60/MARY SARAH, DO/DAUR/-/5/DO/DO, DO  
 60/ANN, DO/DAUR/-/2/-/DO, DO  
 60/SUSAN, DO/DAUR/-/1N/-/DO, DO  
 70  
 50/52/ALMS HOUSE EAST STREET  
 60/MARK, COOK/HEAD/MAR/74/-/PAUPER SHEPHERD/DO, WORTH MATRAVERS  
 60/MAY, DO/WIFE/MAR/-/73/DO/DO, WINTERBOURNE  
 70  
 50/53/DO  
 60/JANE, SERLEY/HEAD/U/-/56/DO/DO, CORFE CASTLE  
 89/((LONG LINE APPEARS TO HAVE BEEN DRAWN UNDER ABOVE BY MISTAKE +  
 THE ENUMERATOR CORRECTED THIS))  
 60/WILLIAM, /SON/MAR((?))/34/-/DO/DO, DO  
 89/((NO DITTO OR SURNAME FOR ABOVE))  
 70  
 50/54/DO  
 60/JOHN, PARKER/HEAD/MAR/44/-/DO (AG LAB)/DO, DO  
 89/((THE ENUMERATOR ORIGINALLY ENTERED HANDLEY AS THE PARISH OF BIRTH BUT +  
 CROSSED IT OUT AND SUBSTITUTED A DITTO))  
 60/ELIZABETH, DO/WIFE/MAR/-/49/DO/DO, HANDLEY  
 89/((SHORT LINEDRAWN UNDER THE ABOVE BY MISTAKE LATER CORRECTED))  
 60/GEORGE, /SON/U/16/-/GENERAL LAB/DO, CORFE CASTLE  
 89/((NO DITTO OR SURNAME FOR ABOVE))  
 50/55/DO  
 60/ELIZABETH, MACLEAN/HEAD/U/-/74/PAUPER/DO, POOLE  
 60/SARAH, STREKLEY((?))/NURSE/W/-/61/DO/DO, WINFRITH  
 70  
 50/56/DO  
 60/HANNAH, GUY/HEAD/MIR/-/28/DO/DO, BURBY IN THE PARISH OF CORFE CASTLE  
 70  
 82/4, /7, 10  
 89/((THE ENUMERATOR ORIGINALLY ENTERED THE NUMBER OF INHABITED HOUSES AS 6 +  
 THEN CHANGED IT TO 5 THEN AGAIN TO 4. ACCORDING TO THE NUMBER OF LONG LINES +  
 IT SHOULD BE 5. ALSO THE WHOLE PAGE WAS BRACKETED IN THE LEFT HAND MARGIN +  
 AND ALONG SIDE IT THE WORD BORO ENTERED))  
 .  
 .  
 90

Figure 2: An Example of the Data Collection File

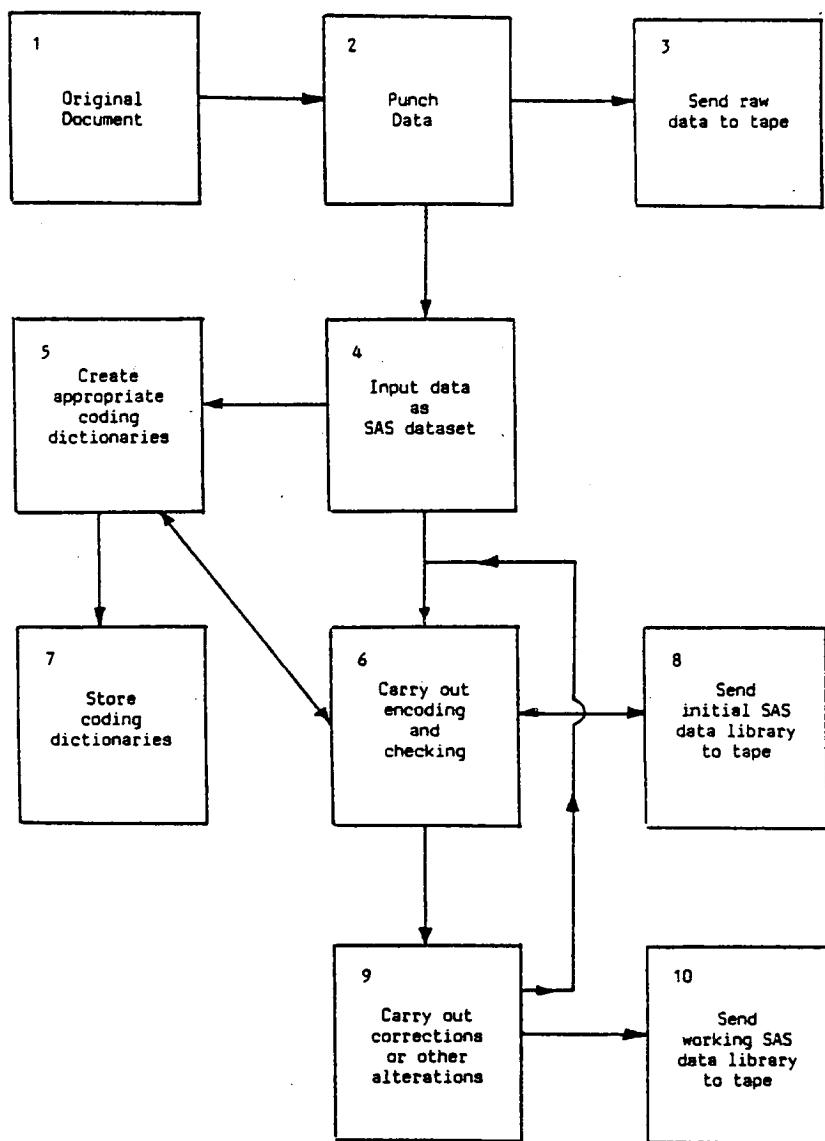


Figure 3: The Input Stage

C A M T A B										P E M F P C M N O L M T R A Y D E R H W									
P	M	O A	H	B	S	R	C	O	D	B	B	P	C	M	T	Y			
1	THOMAS	SCOTT	HEAD	MARR	BREPAD														
2	MARTHA	SCOTT	WIFE	MARR	DRESSMAK														
3	WILLIAM	EDWARDS	HEAD	MARR	BOOT & SHOEMAK														
4	MARY	EDWARDS	WIFE	MARR	BOOT & SHOEMAK														
5	GEORGE	EDWARDS	SON	MARR	BOOT & SHOEMAK														
6	EDWIN	EDWARDS	SON	MARR	SCHOLAR														
7	EDMOND	EDWARDS	SON	MARR	SCHOLAR														
8	JAMES	EDWARDS	SON	MARR	BRICK MAKER														
9	WILLIAM	WILLIAMS	HEAD	MARR															
10	ELIZTH	WILLIAMS	WIFE	MARR	FARM LABOURER														
11	JOHN	MACKINTOSH	HEAD	MARR	FARM LAB NIFE														
12	MARGARET	MACKINTOSH	WIFE	MARR															
13	ELIZTH	MACKINTOSH	DAUGH	MARR															
14	WILLIAM	MACKINTOSH	SON	MARR															
15	MARRAIT	LAMMAS	SEVNT	U	SEVNT														
16	MARRAIT	LAMMAS	SEVNT	U	SEVNT														
17	MARY BELLON	BURPITS	VISITOR	MARR	LABOURER														
18	DAVID	DAVIES	HEAD	MARR	LAB NIFE														
19	ELIZTH	DAVIES	WIFE	MARR	SHOEMAK														
20	JOHN	DAVIES	SON	MARR	SHOEMAK														
21	ANN	DAVIES	DAUGH IN LAW	MARR	SHOEMAK														
22	BERTHA	DAVIES	GRAND DAUGH	MARR	SHOEMAK DAUGH														
23	ELIZTH	ONENS	HEAD	WIDOW															
24	JOHN	ONENS	SON	WIDOW															
25	MARTHA	ONEN	GRAND DAUGH	U	CARTER														
26	GEORGE	ONEN	GRANDSON	G DAUGH	G SON														
27	MARRAIT	ONENS	GRAND DAU	G DAU	G DAU														
28	JOHN	EDWARDS	SEVNT	MARR	SEVNT														
29	RICED	EDGE	HEAD	MARR	BLACKSMITH														
30	ELIZTH	EDGE	WIFE	MARR	BLACKSMITHS WIFE														
31	MARIA	EDGE	DAUGH	MARR	BLACKSMITHS														
32	WILLIAM	EDGE	SON	MARR	SCHOLAR														
33	BARAH	EDGE	DAUGH	MARR	SCHOLAR														
34	JOSE	BENION	HEAD	MARR	LABOURER														
35	MARGARET	BENION	WIFE	MARR	LAB NIFE														
36	JOSE	BENION	SON	MARR	SCHOLAR														
37	MARY	BENION	DAUGH	MARR	SCHOLAR														
38	EDMOND	CANNAS	HEAD	MARR	LABOURER														
39	MARIA	CANNAS	WIFE	MARR	LAB NIFE														
40	EDMOND	CANNAS	SON	MARR	SCHOLAR														

The three variables on the far right H, M, F are automatically generated by the computer. These form a unique identification number to the household, individual and houseful respectively. The variable Personid numbers individuals within the household, sequentially.

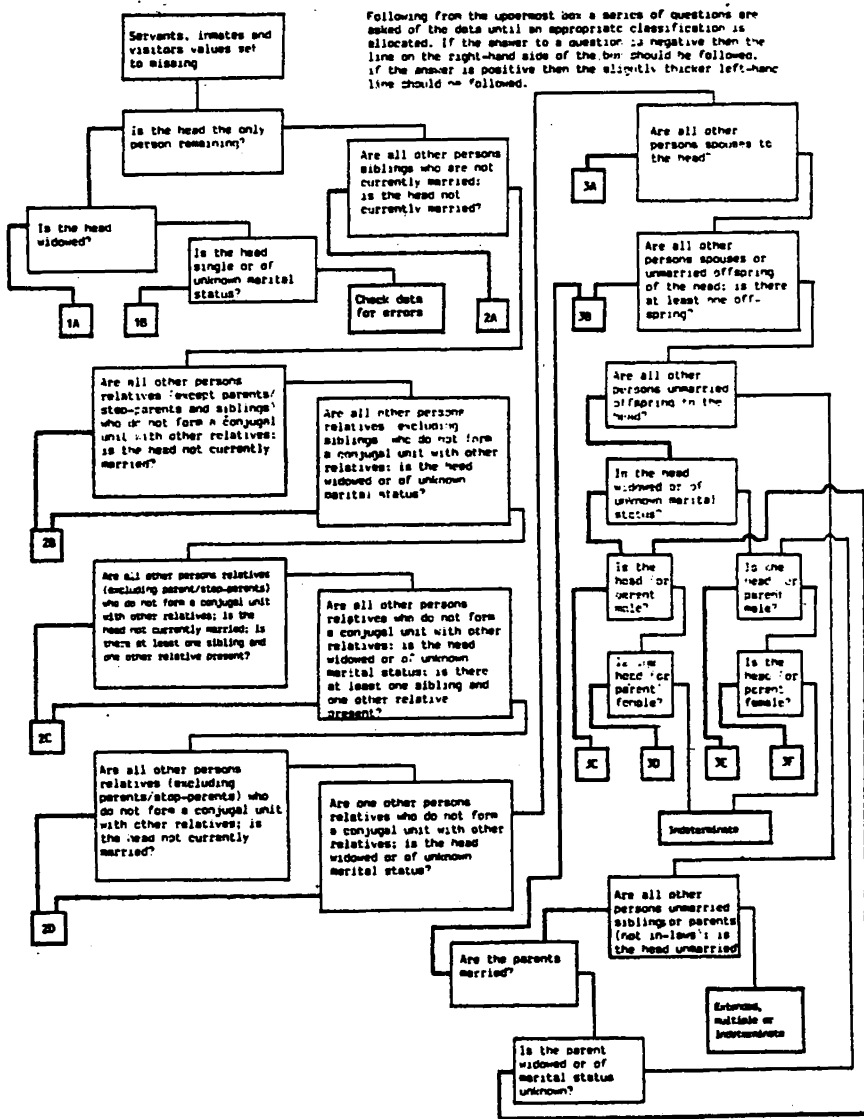
Figure 4: Example of the Data Library Member SPBASE (Broughton, Flintshire - 1851)

TABLE 5.3 HOUSEHOLD STRUCTURE: HOUSEHOLDS BY KIN COMPOSITION

Household types	Classes	Number of households	Proportion of all households
1. <u>Solitaires</u> (singletons in households)	1a) given as widowed 1b) given as non-married or of unknown marital status	_____	_____
	Sub-total	_____	_____
2. <u>No family households</u> (co-residents amongst whom no conjugal family unit can be discerned)	2a) co-resident siblings only 2b) other co-resident relatives only 2c) co-resident siblings and other relative(s) 2d) co-residents with no familial relationship given	_____	_____
	Sub-total	_____	_____
3. <u>Simple family households</u> (conjugal family units only)	3a) married couples without offspring 3b) married couples with offspring 3c) widowers with offspring 3d) widows with offspring 3e) unmarried male parent with illegitimate offspring 3f) unmarried female parent with illegitimate offspring	_____	_____
	Sub-total	_____	_____
4. <u>Extended family units</u> (conjugal family units having kin-linked individuals)	4a) extension upwards (of which ___ fathers ___ mothers) 4b) extension downwards (of which ___ grandchildren only) 4c) extension sideways (of which ___ brothers only ___ sisters only) 4d) combinations of 4a-4c, or any other form of extension	_____	_____
	Sub-total	_____	_____
5. <u>Multiple family households</u> (two or more kin-linked conjugal family units)	5a) households with secondary units disposed upwards (of which ___ also extended) 5b) households with secondary units disposed downwards (of which ___ also extended) 5c) households with secondary units disposed sideways (of which ___ also extended in other directions) 5d) households with secondary units disposed sideways, no member of parental generation present (of which ___ also extended) 5e) combinations of 5a-5d, or any other multiple household arrangement (of which ___ also extended)	_____	_____
	Sub-total	_____	_____
6. <u>Indeterminate</u> (households where kin linkages are insufficient for classification in any of above)		_____	_____
	Sub-total	_____	_____
	TOTAL	_____	100

Definitional and explanatory notes to this table will be found below.  
Researcher's comments to be inserted below.

Figure 5: Model Tables for Analysing Families, Households and Housefuls, prepared by the Cambridge Group



'not currently married' - unmarried; widowed; married spouse absent or of unknown marital status  
 'widowed' - married; spouse absent; separated or divorced  
 Also note that the system allows for polygamous relationships hence the use of the term 'spouses'

Figure 6: Assignment of Household Classification

KONRAD H. JARAUSCH  
Some Reflections on Coding

Coding seems to be one of the necessary evils of historical social research. While there has been much progress with textual data bases or with formalized data entry in non-numerical fashion <sup>1</sup>, most quantitative historical research does transform sources into numerical expressions. The absence of universal standards and the decentralization of historical research have led to a widespread "coding chaos" and to the continual reinvention of the wheel by individual scholars. The waste of time and effort involved has produced periodic appeals for standardization of codes <sup>2</sup>. But in my own area of research (modern German social history) they have not yet had much success.

While some of the coding frustration must stem from the contrariness of human nature, lack of agreement on general codes has deeper reasons. One set of causes involves the divergent institutional foundations of historical research. While on the continent *Grossprojekte* tend to dominate the scene, in Anglo-American countries individual cottage industry seems to be the prevailing mode. The former tends to foster some degree of local uniformity while the latter leads to greater idiosyncrasy. More important for the coding problem is the basic ambivalence of the process itself, which will not simply yield to appeals for good will: On the one hand coding is generalizable and repetitive, dealing with uniform topics across locale and time (such as the distinction between male and female). On the other hand creating a codebook is heavily dependent upon the specification of the research hypothesis and therefore peculiar to each project (such as an investigation of the mobility vs. the working class). This inherent intellectual difficulty will continue to defeat calls for greater uniformity.

Perhaps it would be more practicable to distinguish three different aspects of the coding process, each amenable to a different degree of systematization.

- 1) First there are the technical rules (*Kunstregeln*) which can be found in any suitable textbook<sup>3</sup>. They concern such principles as *Eindeutigkeit*,

---

<sup>1</sup> See the papers by G. Jaritz / A. Müller in this volume on pp. 93ff and E. Mergenthaler: Text Base Management Systeme - Werkzeuge zur Archivierung und Analyse sprachlicher Daten, in: Angewandte Informatik 6 (1983) 262-267.

<sup>2</sup> Cf. the papers by H. Reinke and H. Schultz in this volume on pp. 125ff and 179ff.

<sup>3</sup> M. Thaller: Numerische Datenverarbeitung für Historiker, Wien, 1982, pp.

*Systematik*, completeness, efficiency and *Rückverfolgbarkeit* on which there is bound to be widespread agreement.

- 2) Second, there are many areas in which it is preferable to employ codes developed by historical contemporaries – especially when case studies are later to be compared with aggregate figures available in published sources. The geographical/administrative classifications of the Prussian or German statistical office are a good example. Would it not be more efficient simply to use the numbers for the Prussian provinces and administrative districts (starting with 01 for East Prussia) than to make up a set of one's own, especially if one wanted to compare these findings with data from the relevant *Volkszählung*?
- 3) Third, there are other sizeable fields in which a uniform set of codes would be counterproductive. For instance, in social stratification research the purpose of the investigation determines the structure of the coding scheme. It would be useless to distinguish between different levels of nobility in a study of the Berlin working class of the 19th century. Conversely, it would be nonsense to draw fine lines between manual and non-manual labor in an investigation of the social origin of German students during the last two centuries – since less than 2% stemmed from these strata, whereas it might be interesting to differentiate the dominant *Bildungsbürgertum* into educated officials and free professionals<sup>4</sup>.

Because such problems are inherent in the coding process, three different steps are necessary to reduce the chaos. First, there needs to be more explicit discussion of practical techniques. For instance, the basic principle that coded figures ought to be as diverse and close to the source as possible (whether it be numerical or nominal) is worshipped more often in the breach than in the observance. There is no need to mention individual names (and to embarrass colleagues) lest they turn around and accuse the accusers, but examples of thoughtless and premature aggregation could easily be multiplied. This is where the pilot study can be helpful and where non-numerical data entry is often essential, so that early and sometimes mistaken coding decisions can later be reversed. Second, it seems essential to develop, in those areas where there will be little ideological or methodological disagreement, not only com-

---

126 sqq. and K.H. Jarausch, G. Arminger and M. Thaller: *Quantitative Methoden in der Geschichtswissenschaft: Eine Einführung in die Forschung, Datenverarbeitung und Statistik*, Darmstadt, 1985. See also Floud, Dollar-Jensen or Shorter for earlier examples.

<sup>4</sup> K.H. Jarausch: *Students, Society and Politics in Imperial Germany: The Rise of Academic Illiberalism*, Princeton, 1982, 114ff.

mon codes but also common data sets on tape or diskette. Why does someone using urbanization or place of birth as variables have to look up the size and location of each German village, town or city in some volume of the *Deutsche Statistik*? It seems not only possible but imperative to transform that information into machine-readable form and make it accessible so that the computer can do the matching during data entry (when the name is typed in) – and not the tired researcher or the assistant! Third, coding more controversial and complex variables could be simplified greatly if scholars understood the beauty of multiple codes more widely. Instead of having to argue about the one and only stratification system, why not classify the same datum, namely occupation, according to different principles? For instance one consideration could be Max Weber's conception of power over others (which might divide professions into elite, intermediate and dependent groups). Another might be the economic functional classifications of the German statistical office such as agriculture, commerce, industry, servants, government or professions as well as no-profession. A final consideration might involve social strata – as perceived by contemporaries themselves. This could lead to a stratification scheme starting with the *Bildungs-* and *Besitzbürgertum*, including the *Alte* as well as *Neue Mittelstand* and the working class ... There is no need to argue the specifics of this case developed for research into 19th century German higher education <sup>5</sup>. It is principle that matters. The advantage of a multidimensional approach is not only that it reproduces the complexity of social space better than a one-dimensional scheme, allowing the researcher to try different conceptual dividing lines on the data empirically and letting him thereby determine where the significant distinctions in the data lie. It also makes it possible to use competing sets of classifications which make results more comparable between researchers (if someone else's scheme is also employed).

While the above recommendations themselves do not magically solve all coding problems, they suggest a direction for debate which ought to be more fruitful than endless argument about the various merits and flaws of a single scheme. Clearly, coding is a crucial step in quantitative historical research, since categories tend to predetermine results. As it begins the interpretation process, it will be left to uninformed student assistants only at considerable intellectual peril. If data-sets, whether they be institutional or individual, are not to be buried with each project but be used in secondary analysis,

---

<sup>5</sup> W. Hubbard and K. H. Jarausch: Occupation and Social Structure in Central Europe: Some Notes on Coding Professions, in: *Historical Social Research*, 11 (1979), 10-19.



they must be clearly documented as well as cleanly coded. In some cases exchanging the original documents might be necessary in order to introduce more complex analytical distinctions<sup>6</sup>. But in most instances careful coding techniques, generally agreed upon contemporary schemes and multidimensional classification ought to suffice for secondary analysis and a comparison of results. There is no need for coding over-kill by spending two thirds of the research effort to elaborate a "perfect" code-book. But if quantitative historians are to reap the full benefits of their computer usage, they need to begin cleaning up their coding act!

---

<sup>6</sup> See the secondary analysis of the Kater NSDAP sample for a subsample of professionals, semi-professionals and proto-professionals by this author forthcoming in the *German Studies Review*.

## Probleme sozialökonomischer Klassifikation

Jeder Historiker, der Massenquellen bearbeitet, um sozialökonomische Strukturen und Prozesse zu erhellen, steht vor dem Problem der Klassifikation. Das einzige Merkmal, das die Quellen hierfür einigermaßen regelmäßig bieten, ist die Angabe des Berufes. Dieses Merkmal ist weitaus schwieriger zu verarbeiten als andere für die historisch-demographische Analyse wesentliche, wie das Geschlecht, das Alter, der Familienstand oder der Verwandtschaftsgrad. Die Schwierigkeiten ergeben sich vor allem aus drei Eigentümlichkeiten der Berufsangaben:

1. Die *Vielfalt*. Schon eine Stadt mittlerer Größenordnung weist zwischen 200 und 500 verschiedenen Berufsangaben auf.
2. Die *Abhängigkeit von Raum und Zeit*. Das System der Berufs- oder Standesangaben verändert sich grundlegend mit den Übergängen von einer historischen Epoche bzw. von einer sozialökonomischen Gesellschaftsformation zur anderen, da es den Grad der gesellschaftlichen Arbeitsteilung und die herrschenden Produktionsverhältnisse widerspiegelt. Es ist zudem von regionalen/örtlichen Traditionen abhängig. (Die Bezeichnung "Gärtner" in der Bedeutung Kleinbauer ist z.B. an den sächsischen Raum gebunden.)
3. Die *fehlende Eindeutigkeit*. So sind z.B. "Weber", "Fabrikant", "Angestellter" ohne weitere Erläuterungen und/oder Kenntnis des Umfeldes sozialökonomisch schillernde Begriffe. Hinzu tritt das Problem der Synonyme.

Trotz dieser schwerwiegenden Probleme ist es in der historischen Forschung unbezweifelt, daß Berufsangaben einen brauchbaren Zugang zur sozialökonomischen Struktur der Gesellschaft eröffnen.<sup>1</sup>

Die erste der genannten Schwierigkeiten, die große Vielfalt, schließt eine unmittelbar quantifizierende Analyse auf der Grundlage der Berufe aus. Die statistisch nicht vertretbare Zersplitterung des Datenmaterials, die unüberschaubare Menge der Ergebnisse, der für jede weiterführende theoretische Einsicht viel zu niedrige Abstraktionsgrad, sind offensichtlich. Dies hat jeden Forscher dazu geführt, ein System zur Klassifikation der Berufsangaben zu benutzen bzw. zu entwickeln. Die Zahl dieser Systeme ist sehr groß, fast ebenso groß wie die Zahl entsprechender Untersuchungen. Denn nur selten

---

<sup>1</sup> Hartmut Zwahr: Zur Konstituierung des Proletariats als Klasse. Strukturuntersuchung über das Leipziger Proletariat während der industriellen Revolution, Berlin, 1978, S. 129.

wird ein vorgefundenes Klassifikationsschema übernommen. Eines der wenigen mir bekannten Beispiele ist die Anwendung der von F.G. Dreyfus für Mainz entwickelten Bevölkerungsgliederung durch Etienne François, der sie für Koblenz benutzt hat.<sup>2</sup>

Die Entwicklung jeweils besonderer Klassifikationsschemata resultiert aus den örtlichen Besonderheiten der Quellen und der Sozialstruktur. Ist sie aber nicht auch Ausdruck einer gewissen Überschätzung dieser lokalen und regionalen Besonderheiten durch den Forscher? Vor allem scheint mir die Vielfalt und mangelnde Übereinstimmung der Systeme Ausdruck noch unzureichender Reflexion und Diskussion dieser Frage durch die sozialhistorische Forschung zu sein. Dabei handelt es sich um ein Problem der historischen Typisierung, das wesentliche Voraussetzung für die Theoriebildung auf "mittlerer Ebene" ist.<sup>3</sup> Die Wahl des Klassifikationsschemas entspricht der Annahme eines Modells gesellschaftlicher Strukturen, das die Ergebnisse der quantifizierenden Auswertung der Quellen deutlich beeinflusst und die daraus zu gewinnenden verallgemeinernden Aussagen prägt.<sup>4</sup>

Dies erschwert schon gegenwärtig die Vergleichbarkeit und theoretische Abstraktion mehrerer Untersuchungen. Wenn die einzelnen Untergruppen verschiedener Untersuchungen nicht etwa die gleichen Berufe umfassen, sondern beispielsweise die Tagelöhner einmal mit den selbständigen Fuhrleuten und Schiffern, ein anderes Mal mit Gärtnern und Weinbauern, ein drittes Mal mit ungelernten Fabrik- oder Manufakturarbeitern vereinigen, dann messen sie unterschiedliche Sachverhalte. Der um Theoriebildung bemühte Historiker kann sich dann auch angesichts detaillierter Ergebnisse, die auf Kompartimenten genau ermittelt werden, oft nur noch in allgemeine Aussagen über "Oben", "Mitte" und "Unten" der Gesellschaft retten. Dies ist sicher nicht nur für marxistische Historiker unbefriedigend, denen es um Klassenanalyse geht.

Das Problem scheint mir aber nun um so mehr zu einer weiteren Klärung und Lösung zu drängen, als der Aufbau sozialhistorischer Datenbanken heterogener Herkunft und der Austausch von Datenbanken erwogen werden.

Man kann dem entgegenhalten, daß mit der quellentreuen Berufs- und Standesbezeichnung für jeden Nutzer und Nachnutzer die Möglichkeit gege-

---

<sup>2</sup> Etienne François: Koblenz im 18. Jahrhundert. Zur Sozial- und Bevölkerungsstruktur einer deutschen Residenzstadt, Göttingen, 1982, S. 58.

<sup>3</sup> Wolfgang Küttler: Probleme geschichtswissenschaftlicher Typisierung, in: Zeitschrift für Geschichtswiss., 32. Jg., 1984, H. 12, S. 1070.

<sup>4</sup> Manfred Thaller: Numerische Datenverarbeitung für Historiker, Wien / Köln, 1982, S. 126

ben ist, sein eigenes Klassifikationsschema auf beliebige Datenbanken anzuwenden. Dies ist aber m.E. wegen der unter 2. und 3. genannten Eigenarten von Berufs- bzw. Standesbezeichnungen nicht möglich. Eine hinreichend genaue soziale Qualifizierung der Berufsangaben kann kaum ohne detaillierte Kenntnis des historischen Kontextes der Quelle vorgenommen werden. Zusätzliche systematische Fehler müßten die Ergebnisse verzerren.

Hinzu kommt, daß sich aufgrund dieser Eigenarten die Berufsbezeichnungen der automatischen Klassifikation mit Hilfe der EDV entziehen. Sie müssen zuvor in eindeutige Begriffe transformiert werden, um die grundsätzliche mathematisch-logische Voraussetzung der Quantifizierung zu erfüllen. Dies würde für mehrere Nutzer derselben Datenbank einen großen Aufwand an Mehrfacharbeit erfordern.

Der Ausweg wäre die Klassifikation durch den Forscher, der die Daten erhebt, nach einem Modell, das zwischen allen am Datenaustausch Interessierten vereinbart wird.

Aber ist das überhaupt möglich angesichts unterschiedlicher, ja gegensätzlicher theoretisch-methodologischer Ansätze, angesichts der konträren Zielsetzungen, die insbesondere marxistische und nichtmarxistische Historiker mit der sozialökonomischen Analyse verfolgen?

Der Kern dieser Gegensätze liegt jedoch nicht auf dem Felde der Datenbeschreibung und Datenanalyse, zu dem die sozialökonomische Klassifikation gehört. Er liegt in der Einordnung in die historische Aufeinanderfolge ökonomischer Gesellschaftsformationen und in den weltanschaulich-politischen Schlußfolgerungen. In dieser Hinsicht sind noch immer die bekannten Bemerkungen von Karl Marx in einem Brief an Joseph Weydemeyer erhalten: "Bürgerliche Geschichtsschreiber hatten längst vor mir die historische Entwicklung dieses Kampfes der Klassen, und bürgerliche Ökonomen die ökonomische Anatomie derselben dargestellt. Was ich neu tat, war 1. nachzuweisen, daß die Existenz der Klassen bloß an bestimmte historische Entwicklungsphasen der Produktion gebunden ist; 2. daß der Klassenkampf notwendig zur Diktatur des Proletariats führt; 3. daß diese Diktatur selbst nur den Übergang zur Aufhebung aller Klassen und zu einer klassenlosen Gesellschaft bildet."<sup>5</sup>

Die Trennlinie zwischen marxistisch-leninistischer und nichtmarxistischer Sozialgeschichtsforschung scheidet also nicht unbedingt klassenanalytische Modelle von solchen diverser sozialer Schichtungen. Die Notwendigkeit des Klassenbegriffs insbesondere für die Verknüpfung von "Arbeiterschafts- und Arbeiterbewegungsgeschichte" verfißt daher auch Jürgen Kocka, der gleich-

---

<sup>5</sup> Karl Marx/Friedrich Engels: Werke, Bd. 28, S. 577.

zeitig feststellt: "Der Begriff der Klasse stand und steht nicht gerade im Zentrum der Gunst der bundesdeutschen Sozialhistoriker. Man wählt ihn lieber als Gegenstand begriffshistorischer Untersuchungen statt als tragendes Instrument der eigenen Analyse."<sup>6</sup>

Da aber die Klassen objektiv in der Gesellschaft existieren, spiegeln sozialökonomische Klassifizierungssysteme bürgerlicher deutscher Sozialgeschichtsforschung von Otto Konrad Roller bis Hartmut Kaelble diese Struktur mehr oder minder deutlich und konsequent wider. Dies geschieht allerdings nicht auf der Ebene der Klassen der feudalen bzw. der kapitalistischen Gesellschaft, sondern auf der Ebene sozialer Gruppen, die Strukturelemente der Klassen bilden. Hier ist ein relativ großes Maß an Übereinstimmung aller bisher verwandten Klassifizierungsschemata festzustellen.

Für die Zwecke der Datenbeschreibung, des Aufbaus internationaler Datenbanken und des Datenaustauschs ist eine Klassifikation der Berufsbezeichnungen auf der Basis sozialer Gruppen das Gegebene. Diese Klassifikationsebene ist nicht nur hinreichend, sondern notwendig für die Analyse der inneren Struktur von Klassen, für die Untersuchung von Klassenbildungs- und -zerfallsprozessen. Ließe man diese Ebene außer acht und nähme die Klassifikation unmittelbar nach Klassen und Schichten vor, wäre deren Binnenstruktur verdeckt und die Analyse der Entwicklungsprozesse kaum noch möglich. Andererseits setzt die Klassenanalyse voraus, daß die sozialen Gruppen so gebildet werden, daß ihre Zusammenfassung zu den Klassen und Schichten der jeweiligen Gesellschaft möglich ist. Sie sollten also Klassengrenzen möglichst nicht übergreifen und damit verwischen. Man kann das Verhältnis von Klassen/Schichten einerseits und sozialen Gruppen andererseits etwa so fassen: Die sozialen Gruppen sind von Besonderheiten der sozialen, ökonomischen und politischen Struktur im Verhältnis zu den allgemeinen Wesensmerkmalen der Klassen und Schichten geprägt.

Um dies zu verdeutlichen, sei ein Klassifikationsmodell vorgestellt, das bei einer Untersuchung der Sozialgeschichte Berlins im Manufakturzeitalter entwickelt wurde und inzwischen auch bei der Erforschung anderer Städte und Gewerbeorte in der Übergangsepoche vom Feudalismus zum Kapitalismus erprobt wird:

---

<sup>6</sup> Jürgen Kocka: Klassen oder Kultur? Durchbrüche und Sackgassen in der Arbeitergeschichte, in: Merkur, Dt. Zs. f. europäisches Denken, Jg. 36, 1982, H. 10, S. 960.

1. *Feudalklasse*

- (11) – Landadel
- (12) – Hof- und Beamtenadel
- (13) – Offiziere
- (14) – Hohe bürgerliche Beamte und hohe Geistlichkeit

2. *Handels- und Manufakturbourgeoisie:*

- (21) – Kaufleute und Bankiers
- (22) – Manufakturunternehmer

3. *Intelligenz und Beamte*

- (31) – Akademiker (Pfarrer, Ärzte, Advokaten etc.)
- (32) – Künstler und Techniker
- (33) – Mittlere Beamte
- (34) – subalterne Beamte

4. *Einfache Warenproduzenten und Kleingewerbetreibende:*

- (41) – Kleinhändler
- (42) – Gastwirte und Transportgewerbe
- (43) – Zunftmeister
- (44) – unzünftige selbständige Handwerker
- (45) – städtische Agrargewerbe

5. *Lohnarbeiter*

- (51) – Gesellen
- (52) – Manufakturarbeiter und verlegte Handwerker
- (53) – Tagelöhner
- (54) – Diensthöten
- (55) – Arme, Insassen von Zucht- und Arbeitshäusern

6. *Soldaten*

- (61) – Unteroffiziere, Feldstern und Spielleute
- (62) – Soldaten

7. *Landbevölkerung*

- (71) – Bauern
- (72) – Kleinstellenbesitzer
- (73) – Landhandwerker und ländliche Gewerbetreibende
- (74) – Landarmut

9. *Unsichere und fehlende Angaben:*

- (91) – Keine Angabe
- (92) – Bürger und Eigentümer
- (93) – Kolonisten, Immigranten ohne Berufsangabe
- (94) – Witwen ohne Berufsangabe des Mannes
- (95) – Juden ohne Berufsangabe

Dieses Klassifikationsschema ist alles andere als vollkommen. Die Klassensynthese in den Zehnergruppen ist problematisch. Kann, soll man die bürgerlichen Beamten wegen ihrer Herrschaftsfunktion als Teil der Feudal-klasse begreifen? Sind die Soldaten tatsächlich eine eigene Schicht der spätfeudalen Gesellschaft, oder sind sie insbesondere unter den Bedingungen des preußischen Freiwächtersystems nicht eher den Lohnarbeitern zuzuordnen? Die 70er Gruppen der Landbevölkerung sind im 17./18. Jh. schon längst keine einheitliche Klasse mehr. Und für die fehlenden Werte der 90er Gruppen ist ein solcher Begriff ohnehin unsinnig. Aber die Erkenntnis der Klassenstruktur ist im Verlauf der Untersuchung zu vervollkommen. In diesem Zusammenhang geht es um das Klassifikationssystem der sozialen Gruppen. Auch da gibt es zahlreiche Probleme, die in der praktischen Arbeit der Datenerfassung erhebliche Schwierigkeiten brachten und zu Kompromissen führten: beispielsweise die Trennung von Kaufleuten und Manufakturunternehmen, von zünftigen und unzünftigen Handwerkern, von selbständigen und verlegten Handwerkern. Wir haben wegen der allzu großen Unsicherheiten auf diese Unterscheidungen schließlich verzichtet. Ein verwandtes Problem ist das der Doppelangaben von Berufen ('Soldat und Maurergeselle'). Die Klassifikation zwingt zur Entscheidung (für die jeweils erste weil derzeit bestimmende Tätigkeitsangabe?) und zum Verzicht auf ein Stück realer gesellschaftlicher Vielfalt.

Es zeigt sich, daß dieses Klassifikationssystem wie jedes andere die sozialökonomische Mehrdeutigkeit von Quellenbezeichnungen zugunsten einer Eindeutigkeit aufhebt, die nicht voll der Realität entspricht, sondern eine Abstraktion der historischen Wirklichkeit ist. Die Übergangsfelder zwischen sozialen Gruppen werden zu Grenzen. Eine solche Reduktion der gesellschaftlichen Wirklichkeit liegt im Wesen theoretischer Verallgemeinerung, die jedes Klassifikationssystem darstellt. Das bekannte Reliabilitätsproblem ist hiermit verknüpft. Die Unsicherheiten der Zuordnung lassen sich mit einer wachsenden Zahl sozialer Gruppen nicht vermindern, sondern sie nehmen notwendig zu, da die Anzahl der Grenzziehungen vermehrt wird. Die Diskussion um möglichst allgemeingültige Klassifikationssysteme sollte also auf eine einfache Struktur zielen, auf eine optimale Anzahl von sozialen Gruppen, die nach unserer Erfahrung größer als 10 und kleiner als 30 sein muß.

Das vorgelegte Klassifikationssystem ist ferner alles andere als universell. Es ist auf die Analyse städtischer Gesellschaft in der Epoche des Übergangs vom Feudalismus zum Kapitalismus zugeschnitten und an mitteleuropäischen Verhältnissen orientiert. In einer internationalen Diskussion ließe sich zweifellos ein höheres Maß an Universalität erreichen. Aber es wird die Notwen-

digkeit bleiben, für unterschiedliche Gesellschaftsformationen jeweils eigene Klassifikationssysteme zu entwickeln.

Insgesamt erscheinen die Schwierigkeiten einer Konvention über Klassifikationssysteme zur sozialökonomischen Analyse von Massendaten außerordentlich groß. Eine Bewältigung dieser Probleme, oder erst einmal eine Verständigung über diese Fragen wäre jedoch als Voraussetzung für die Schaffung internationaler Datenbanksysteme wichtig. Diese Ausführungen möchten ein Diskussionsbeitrag dazu sein.





## **A Data Base for Undergraduates on a Micro**

For the past three terms we have been working on a study of immigrants in Winnebago County (Wisconsin) for the period 1906 to 1916. Our study began with a 1982 University of Wisconsin-Oshkosh Faculty Development Grant to create an experimental history course in which students could learn to use computers for quantitative history. The course has been offered for two terms and the students have discovered that the records are voluminous and the work time-consuming, for in their research they have used immigration and naturalization records of Winnebago County, federal and state censuses, city directories, probate court records, marriage licenses, death certifications, register of deeds records, plat records and maps, applications for employment with local industries, church records, records of social groups, and iconographic records. Thus the students were able to trace the immigrants and their families from the immigration and naturalization records into the community, and in quite a number of cases up to the present day.

For this project the data was entered on the PDP 11/70 digital computer. An indexed sequential file was created. The program, written in BASIC-PLUS, prompts the user to enter data that has been previously gathered. Data is checked for character length, data type (numeric, character) and ranges (day, month, year). These checks are made for each response to a question; for example, country born? If the user answers with more than twenty characters, an appropriate error response is displayed and that question is displayed again.

### **The Project**

Once the data has been entered for immigrants, the student may print this data, display it back to the CRT, or enter the update section of the program and change any mistakes made on individual questions. After the errors have been corrected a program is run to extract the data from the indexed file and create a new sequential file compatible with SPSS-X (a statistical software package). In addition, a couple of small programs are used to change the file format of the data. The data is then backed off to tape as a sequential file, but may be copied back to disk and converted to an indexed file for updating.

The statistical package SPSS-X is set up using FILE TYPE GROUPED in order to handle multiple and variable numbers of records per case; variable only because there are three records per child and up to fifteen children allowed. A fixed length record could have been created during the input program

if the maximum number of records would have been created and written to the file with all variables being set to missing in the superfluous records for families with less than 15 children, but this would have substantially increased the file size. Indeed, we soon discovered that the data collected per case was huge. For example, for the primary immigrant there were 19 records (of 80 bytes each) per case; for the child were 3 records per child, with the possibility of 19 children per case, leading to 57 possible records. This resulted in a range of a minimum of 19 records per case to a maximum of 76 records per case, with a maximum of 4608 bytes per case.

Inasmuch as several of our students wanted to work on this project during the summer months when the University's computers (PDP 11/70's and VAX 11/780's) are down at various times for maintenance and software installations in preparation for the fall semester, we decided to shift segments of the project to microcomputers. Thus students could obtain reports on the immigrant's age at the time of embarkation and reports on the immigrant's age at the time of petition. These reports show frequencies and percents of every age; and a frequency histogram was created as well showing the mode, mean, median, minimum and maximum. In addition, students could obtain reports on the immigrant's occupation at the time of petition. Thus they could list alphabetically all immigrants, all occupations with their frequencies and percents. They could also obtain complete lists of each immigrant, case number and occupation.

Since the data file is large and contains data not needed for this project, a utility program was written to extract only those fields necessary. As indicated previously, the program written in BASIC PLUS on a PDP 11 reads the first record of the sequential data file and uses the case ID as the key field, the number of children as a secondary field, as that will tell students how many records there are for this case. The data needed is then written to a new sequential data file that is compatible with SPSS-X on the IBM XT. The extracted data is then downloaded to the XT. This was done using an IBM XT with a modem, a phone line to connect with the PDP 11 and a communications package called PCTALK which will transfer the data from the PDP-11 disk to the XT's floppy or hard disk. The data file needs to be on the XT's hard disk prior to running the SPSS-X program. Once this is done, SPSS-X, which is on a floppy disk, is initiated, putting the students into interactive mode with SPSS-X. This gives the students entering commands to analyze age and occupation the benefit of receiving immediately syntax error checks for each command entered. This was not possible on the PDP 11/70 or VAX 11/780. Jobs on the VAX had to be run in batch (not interactive).

I will discuss the 550 immigrant records that the students downloaded

this summer by looking at the results under the following rubrics: Arrival, Petition for Citizenship, Age, Physical Appearance, Occupation, Residency, Marital Status, Family Size, Time Lapse between Petition and Citizenship, Certification, and Rejections.

## Arrival

68 per cent of the 550 immigrants to Winnebago County who are the subject of this study arrived at the port of *New York* (372 people). Narrative accounts of the Ellis Island experience of these New York arrivals are lacking, and statistics cannot provide information about whether the bureaucratic procedures in the vast hall at that point of the debarkation made Ellis Island a purgatory, an *isle of tears*, or a *golden door* for those who eventually came to live in a small city in Wisconsin.

*Baltimore* was the port of arrival for the second largest number (17.5% of the total); *Philadelphia* (23 people), *Boston* (16) and *Portland* (4) were the first stops for a few others.

46 of the 550 arrived first in *Canada*, not at a United States Atlantic port, and they eventually came *across the border at a Great Lakes port*. *Detroit* was the point of entry for *half* of those who traveled this route, with *Buffalo*, *Chicago* and *Milwaukee* each receiving three or four. Of these immigrants from the north, one, James Harra (#388)<sup>1</sup>, an unmarried laborer from Liverpool, England, lived in Quebec for seven years and then took the Grand Trunk Railroad to Detroit. Martin Dolven (#322), a Norwegian railroad worker, followed that same route, as did Hans Rasmussen (#329), an unmarried Danish cabinet maker. Others came by train to Detroit from Halifax, Nova Scotia. One came into Vermont from Cornwall, Canada.

## From Arrival to Petition for Citizenship

The naturalization records provide no detailed information about the 550 petitioners' travels, work, residences, and other experiences between the time they first arrived in America and eventually came into Winnebago County. 107 had made the decision to declare their intention to become citizens before they arrived in Winnebago and had started the naturalization process in

---

<sup>1</sup> Examples of particular immigrants cited in the text are followed by a case number in parenthesis for reference to the manuscript Immigration and Naturalization Records in the Area Research Center of the University of Wisconsin-Oshkosh Polk Library and to the computer analyses which identify each immigrant by the case number from the manuscripts.

another place (37 outside Wisconsin). Gustave Krause (#40), a laborer from Germany, made his declaration in Marquette, Michigan. Jacob Kratochvil (#101), an Austrian tailor, started his naturalization procedures in Cook County, Illinois and his oldest child was born in Chicago, a double indication of prior residence there.

Examination of the *birthplaces of immigrants' children born after the date of arrival in America and before the date of petition* from Wisconsin yields some information about earlier residences or travels. Michael Kren (#34), a farmer from Germany with a German wife, had one child born in Canada and one in North Dakota.

Such bits of information suggest that a few petitioners wandered for a while before reaching Winnebago County. Most others, however, reported their date of arrival at a *United States Atlantic port* as only *two or three days earlier* than their *date of arrival in Wisconsin*, suggesting that they had relatives, friends, contacts or other information about the state perhaps even before they left Europe, bought their train tickets at Ellis Island, and came directly west.

## The Petitioners

The bulk of the data used for this computer analysis relates to the immigrants' condition and status at the time of their petitioning for citizenship in Winnebago County.

### Age

Of the 542 immigrants who reported a birthdate on the naturalization forms, the *majority were in middle age* when they petitioned for citizenship. 166 were in *their 40s*, 150 in *their 30s*. The *mode* in this analysis was age 33. The *mean and media* were age 45. Thus for the *typical* petitioner, about *fourteen years elapsed between the embarkation for America at age 19 and the petition for citizenship at age 33*. At the extremes, the *oldest petitioner* was 79, the *youngest* 23. Ivan Jenkyns (#250), an unmarried retired farmer, had left his native England at age 19, farmed in Wisconsin for most of his life, and received his certificate of citizenship when he was nearly 80 years old. Ernst Wuthrich (#497), an unmarried farmer from Switzerland, migrated when he was 8 years old and became a *citizen at age 23*.

## Physical Appearance

Some of the official naturalization forms included information about the physical appearance of applicants, and a few generalizations can be made from analyzing that data. The men were slight; most were well under six feet tall and weighed about 135 pounds. More interesting was the common recurrence of a missing index finger. Such a mutilation might have happened in sawmills or in other work-related accidents, or it might have been self-inflicted – a draft evasion technique that was not unknown in nineteenth century European countries with compulsory military service.

## Occupation

At the national level, the major labor issue of the new immigration after 1880 was the preponderance of unskilled workers who flooded the labor market, lowered wages or stabilized them at low levels, deterred unionization, and built much of America's transportation and industrial system. *That pattern seems to have prevailed in Winnebago County.* "Laborer" was the most frequent answer to the question in the naturalization papers about the petitioners' occupations. 233, or nearly 37% of the 550 petitioners defined their work that way.

By contrast, 136 people or 25% listed a skilled craft as their means of livelihood. Within that artisan group, woodworkers outnumbered all others. Oshkosh, the urban center of Winnebago County, was known in the early twentieth century as "*Sawdust City*" because of the number of its lumber and woodworking establishments. Perhaps that reputation attracted workers in wood or, more likely, job opportunities and apprenticeships for newly arrived Europeans were most abundant in wood businesses. 21 petitioners listed themselves as *carpenters*, the largest category of skilled workers in the study group. Cabinet maker, contractor, architect-builder, wood turner, timber estimator, lumber inspector were given as the occupations of 23 others.

Besides wood products, *paper*, *cheese*, and *beer* were also produced in Winnebago County and environs; and sixteen papermakers, six cheesemakers, and two brewers were counted among the petitioners. There were also 11 blacksmiths, 7 machinists, 6 painters, 9 tailors, 7 shoemakers, 6 butchers, and one or two representatives of other crafts and trades among them: plumber, mason, bricklayer, iron moulder, cooper, glassblower, engineer, tinsmith, electrician, printer, upholsterer, baker.

*Farmers* made up a little over 15% of the petitioners (83 people). Another 10% were in jobs associated with business, sales, or trade. The *tavern* business was the most popular enterprise for immigrants in Oshkosh; among

the petitioners were 13 saloonkeepers, bartenders, and bar clerks. There were also 11 merchants, 9 peddlers, and 4 traveling salesmen. There were 5 grocers and a sprinkling in other assorted sales jobs: junk dealer, real estate, fish dealer, wholesale school supplies, shoe sales, hardware. There were two manufacturers and one superintendent of a paper company.

A few petitioners had service jobs: 6 firemen, 2 policemen, a parole agent and the register of deeds represented government services. Two barbers, a bootblack, a gardener, an undertaker, and two janitors were in the records. A drayman, a chauffeur, a boatman, a bicycle repairman, three coachmen, and a railroad flagman and section foreman gave some representation to transportation services.

Only 6 of the group were in professions: 2 physicians, 2 male nurses, a chiropractor, and a pastor. Nine had white collar jobs: 8 clerks and a bookkeeper. Completing the list, there were 9 retired persons, 2 students, and an artist.

## Residence

A study of residence patterns among the petitioners yielded some data suggesting that Winnebago County petitioners tended to follow the national pattern of ethnic groups settling together in neighborhoods. As their occupations might suggest, most of the petitioners settled in the city, not on rural routes or in villages. 249 of the 550 (about 45%) lived in Oshkosh, the county seat and largest city in Winnebago County. An additional 170 lived in the two other industrialized urban centers in the county, 99 in Menasha and 71 in Neenah.

The city of Oshkosh in the nineteenth century was divided socially as well as geographically by the Fox River. South of the river was a region of mills and workers' quarters with lower status schools and churches than north of the river. The south side was geographically smaller than the north, but 100 of the 249 immigrant petitioners who lived in the city had residences there.

Only about 20 immigrants lived at the same address, most of them relatives or natives of the same village, but some were apparently boarders. Conrad Schuhart (#154), a laborer from Poljana, Russia with a Russian wife and five children lived at the same residence with Johannes Dippel (#181) a Russian laborer from the same village and a family man with a Russian wife and two children.

John Madison (#262), a Danish woodworker who was a widower without children lived at the same residence with two other immigrants, John Johnson

(#264). a Swedish laborer who was single, and Carl Wahlgren (#263), a Swedish laborer who had a Swedish wife and four children.

*58 of the petitioners lived on rural routes scattered throughout the sixteen townships of the county, and the remainder lived in the small villages of those townships—communities such as Omro, Winneconne, Eureka, Rush Lake, Allenville, and Larsen.*

### Marital Status

The great majority of immigrants seeking citizenship were *married*, 468 of the study group, or nearly 86%. 3 were *divorced*, 13 were widowers, and 62 were unmarried.

*Four of the wives of the married group were listed as living in Europe rather than at their husbands' residences in Winnebago County. Reinhold Jungbauer (#255), a laborer married to a woman, born as he was in Obermoldau, Austria Hungary, listed her residence as unknown; his wife had perhaps not migrated or had returned to Europe, or possibly she had separated from her husband and lived elsewhere in the United States. Such a separation occurred in the marriage of Rudolph Kopiski (#390), a German carpenter, whose Wisconsin-born wife lived in Chicago.*

Nearly one-third of the wives of the petitioners were *born in the United States* (151 wives, or about 31% of the total). Intermarriage across ethnic and national lines occurred most frequently among Canadian immigrants, 73% of whom married women born in the United States. English (67%), Swiss (67%), and Irish (64%) also tended to marry American women more frequently than women of their own nation of origin.

By contrast, 88% of the Russian immigrants married Russian women, and 83% of the Austrians married wives from their own country. 19 immigrants married wives not only from their own nation of origin but also from *their own home town or village*.

61 per cent of the *German* immigrants married German wives, 32% married women born in the United States, one married a Swiss and two married Russians. Scandinavians tended to seek out wives of northern European or American origin. Forty-four per cent of Swedish men married Swedes and the others married wives from Norway, Finland, Germany, and the United States. 52 per cent of Danes married Danes and the others married Norwegians, Swedes, Germans, and Americans. 31 per cent of the Norwegians married wives from their own country, and the others married women from Denmark, Iceland, and the United States.



## Family Size

Of the 468 married petitioners, the great majority, 427, *had children*. The average number of children was 4. The families ranged in size from those with 1 child to one with 14. Fred Galau (#491), a farmer from Germany migrated when he was 17, married a Swiss wife, had his first child when he was 26 and his fourteenth child when he was 51, nine boys and five girls all born in Wisconsin.

Eighty-two families of the 427 households in the study group had more than 7 children: twenty families had 8 children, eighteen had 9, twelve had 10, two had 11, one had 12, and one had 14. (If these figures can be used for social analysis, 10 offspring would seem to have been the acceptable limit on family size, with a dramatic drop between the number of families with 10 and the number with 11.)

The sex division of the children conformed fairly closely to the national average, with a total of 914 female and 836 male children, 78 more female than male. In a few cases, the families produced a string of one sex, had one of the other and then stopped reproducing. John Schneider (#75), a Prussian laborer, and his wife, who was born in Saxony, had eleven girls before John, Jr., the twelfth and last child was born. All the children were Wisconsin natives.

Four families had twins. Two families adopted their children. Adam Block (#9), a Russian laborer married to a Russian wife, had one child, an adopted son named Adam Block, Jr.

## Time Lapse between Petition and Citizenship

The time lapse between the date of petition for citizenship and the date of the certificate of citizenship was about 5 *months* for Winnebago County immigrants. Forty-seven people waited 153 days (5 months and 1 day); thirty-five waited 99 days (3 months, a week and a day). One person had a delay of only 2 weeks; one at the other extreme had a lapse of nearly 8 years. The mean for this analysis was 127.47 days or 4 months and 6 days; the median was 112 days or 3 months and 3 weeks.

## Certification

Certificates of citizenship were issued on a limited number of court days during the year, and the number and date of those days varied somewhat between 1907 and 1913. In 1907, there were 6 certification dates, 8 the next year, 16 in 1909, 10 in 1910, 9 in 1911, 6 in both 1912 and 1913.

The changes in the number of certification dates do not correspond to the changing number of aliens admitted to citizenship during the years after 1907. Only 10 people became citizens in Winnebago County in 1907, and 17 in 1908. There was a sharp increase to 56 in 1909 and 67 in 1910. There was a slight drop in 1911 to 63, but the number went up to 72 in 1912 and to 155 in 1913.

### The Rejected Few

The naturalization process went smoothly for the vast majority of petitioners, but a few encountered some difficulties. Four had problems proving that they had satisfied the United States residence requirements. Otto Wishnowsky (#82), an unmarried carpenter from Danzig, Germany petitioned in 1909, claiming that he had resided in Memphis, Tennessee for a while before 1907; he was unable to prove that, however, so this petition was denied. He petitioned again in 1910 and was again turned down. In the case of John Hewitt (#216), an Ontario machinist, married, with an infant child, the court clerk sent two official requests to Minnesota for an affidavit to confirm Hewitt's claim that he had resided in that state from 1908 to 1910, but neither request was answered. As a result, Hewitt's petition was denied. Edwin Huband (#243), a papermaker from London, had not lived in Wisconsin for the statutory period, so he was unsuccessful in his petition.

Problems with witnesses caused rejections for still other petitioners. On the day before New Year's Eve in 1911, six Russian Jewish immigrants, four of them peddlers, one a shoemaker, and one an oil merchant, all married with families, were rejected because a rabbi, one of the two witnesses used by all of them, was not a citizen. Three of the six - all peddlers - petitioned again with a different witness and secured certificates of citizenship.

Chris Peterson (#148), a blacksmith from Copenhagen, Denmark, had his wife testify as a witness for him, but she was not a citizen so his petition was rejected. In three other cases, however, women served as witnesses, and the petitions were approved without question.<sup>2</sup> In all, 10 petitions were denied because a witness was not a citizen.

---

<sup>2</sup> Peter Oskar (#375), a cement contractor from Schleswig Holstein, married to a Danish woman, had his minister and the minister's wife testify as his witnesses. George Morasch (#549), a Russian laborer with a Russian wife and nine children, had a laborer and his wife as witnesses. John Dunne (#387), an Irish farmer married to an Oshkosh native, had as one of his witnesses Margaret Kelly, a housewife.

Raimund Hoffman (#242), a Bohemian laborer, was turned down because one of his witnesses had not known him for the full period of his Wisconsin residence. Charles Snyder (#93), a painter from Prussia and a Civil War veteran, ran into an objection that one of his witnesses had not known him for five years; that objection, however, was overruled by the court and Snyder secured his certificate of citizenship without further difficulty.

Two others, John Hintzki (#140), a bartender from Danzig, and Adolf Otto (#142), a German carpenter, used witnesses who were unable to prove their acquaintance with the petitioners.

Frank and Joseph Quella (#446, #447), brothers from Germany, one a contractor and one a laborer, used as witnesses a manufacturer and the city treasurer, one of whom, unidentified, was declared to be not qualified to testify. Soren Rasmussen (#195), A Danish cement contractor, used as his witnesses, a contractor and an insurance man and their testimony was deemed "unsatisfactory." All three of these petitioners suffered rejection because of the inadequacy of their witnesses.

Both witnesses for George Wacker (#377), a male nurse from Germany, were male nurses in Oshkosh and one of them failed to appear to testify; Wacker did not secure citizenship as a result. In other such cases of non-appearing witnesses, though, the petition was continued until a later date, or a substitute witness was used. A few petitioners came prepared with three or four witnesses rather than the two required by law.

Problems with the filing of the Declaration of Intent (filed outside the clerk's office or otherwise not properly executed) caused John Gavin (#33), an Irish timber estimator, John Jungbauer (#105), an Austrian laborer, and Frank Binder (#106), an Austrian carriage builder, to have their petitions denied.

Three petitions were dismissed because the petitioner himself did not appear. Boniface Ganther (#135), a mason from Germany, was hospitalized on the appointed day for his appearance and his case was continued. John Sande (#25), a forty-six year old Norwegian shoemaker, a widower with two children, failed to appear, with no explanation given in the records, and his petition was voided. Jens Jensen (#510), a Danish laborer married to a Danish wife, apparently changed his mind about staying in America, returned to his nation of origin, and did not appear on the day assigned to execute his papers. He was the only one of 550 petitioners who voluntarily abandoned the naturalization process and went back home.

Kazmir (Kermien) Pawlowski (#299), a German laborer married to a Wisconsin native, was denied because he was "unwilling" to learn the laws and Constitution of the United States. The suggestion that he chose not to

learn rather than that he was unable to learn because of a language problem distinguished him from 17 other petitioners who secured citizenship even though they were illiterate or semiliterate or spoke very little English.

Of the 550 people who petitioned for U.S. citizenship in Winnebago County between 1906 and 1916, a total of 36 or 6 1/2 % were denied.

Of the remaining 514 who were successful petitioners, 3 had their citizenship revoked after having secured it. Lucian Koschnik (#192), a German tailor married to a German woman was sixty-three years old when the United States entered World War I; he had his certificate of citizenship cancelled by the United States District Court of Eastern Wisconsin on 1 December 1917 with no reason given in the naturalization papers.

August Borschke (#149), a German carpenter married to an Oshkosh native and the father of five children, had his citizenship cancelled on 6 April 1930 by the United States District Court of Eastern Wisconsin.

Hans Jensen (#136), a ninety-three year old unmarried Danish farmer, lost his citizenship and that information was communicated to the Winnebago County Clerk of Court on 23 July 1957 by the Justice Department.

## Conclusion

When the computer analysis of the naturalization records is completed and new students in this continuing research project have time to pursue the personal and family histories of more of the new immigrants in Winnebago County, the cultural and social role of these new citizens as community builders can be fully explored. Such a contribution to local and ethnic history, made in large part by undergraduate students, can confirm for many of us the value of active history for the enrichment of historical studies.



**Von der Quelle zum Datensatz.  
Vorschläge zum Einsatz datenbankorientierter Verfahren  
in der universitären Lehre.**

**1. Vorbemerkung**

Während die grundlegenden Fragen, wie quantifizierende Geschichte an den Universitäten zu lehren wäre, in den späten Siebzigerjahren in der Bundesrepublik Deutschland und auch in Österreich ausführlich diskutiert und zum Gutteil beantwortet wurden,<sup>1</sup> fehlen bislang Konzepte zum Einsatz der quellennahen oder quellenorientierten Datenverarbeitung<sup>2</sup> in der historisch/akademischen Ausbildung. Vorteile und Nutzen dieses zumindest auf dem Gebiet der Geschichtsforschung relativ jungen Zweiges der EDV-Anwendung stehen heute wohl außer Zweifel<sup>3</sup>; so schien es sowohl aufgrund mehrerer in Graz laufender Projekte<sup>4</sup> als auch in Hinblick auf die Verbes-

---

<sup>1</sup> Zu grundlegenden Problemen der Didaktik von quantifizierenden Methoden vgl. *Thomas Schuler*: Quantifizierung und Geschichtsstudium, in: HSR 17, 1981, S. 61f. - *Heinrich Best und Wilhelm H. Schröder*: Basiscurriculum für eine quantitative historische Sozialforschung. Vorschläge für eine Einführungsveranstaltung am Beispiel des ZENTRUM-Herbstseminars, in: HSR 17, 1981, S. 3-50. - *Henning Bauer*: Der Einsatz archivierter Daten in der Lehre der Historischen Sozialforschung, in: HSR 24, 1982, S. 63-72.

Davon zeugen auch die im deutschen Sprachraum erschienenen Lehrbücher, wie *Norbert Ohler*: Quantitative Methoden für Historiker, München, 1980., oder *Manfred Thaller*: Numerische Datenverarbeitung für Historiker, Wien / Köln, 1982 (= Materialien zur Historischen Sozialwissenschaft. 1.).

<sup>2</sup> Vgl. *Konrad H. Jarausch, Gerhard Arminger, Manfred Thaller*: Quantitative Methoden in der Geschichtswissenschaft. Eine Einführung in die Forschung, Datenverarbeitung und Statistik, Darmstadt, 1985, S. 65-69.

<sup>3</sup> Eine Beschreibung der wichtigsten sechs Punkte findet sich bei *Konrad H. Jarausch, Gerhard Arminger, Manfred Thaller*: Quantitative Methoden in der Geschichtswissenschaft. Eine Einführung in die Forschung, Datenverarbeitung und Statistik, Darmstadt, 1985, S. 68.

<sup>4</sup> Neben dem in diesem Band vorgestellten Unternehmen zur Migrationsgeschichte steht in Graz das Projekt "Untersuchung der Einsatzmöglichkeiten der EDV für historische Faktendokumentation im Bereich der historischen Grundwissenschaften (HIFADOK)" in Bearbeitung, das sich dzt. in der Hauptsache mit maschinengestützter Edition und dem Aufbau einer prosopographischen Datenbank

serung und Vorbereitung statistischer Analysen wünschenswert, eine Einführung auf diesem Gebiet in das Lehrangebot aufzunehmen. Eine Möglichkeit der Vermittlung solcher quellennaher und datenbankorientierter Verfahren, wie sie in Graz verwirklicht wurde <sup>5</sup>, soll im folgenden diskutiert werden.

Bereits seit 1980 werden am Institut für Geschichte Lehrveranstaltungen angeboten, die eine Einführung in die Probleme der Quantifizierung zum Inhalt haben <sup>6</sup>. Diese von St. Karner betreuten Übungen betreffen besonders drei große Sachgebiete:

- Vorstellung des EDV-Zentrums und Erlernen der wichtigsten Betriebssystem- und Editorbefehle der UNIVAC 1100.
- Einführung in die Quantifizierung und Diskussion theoretischer und methodischer Fragen.
- Erste Arbeiten mit dem statistischen Programmsystem BMDP.

Die Studierenden, welche die Mühe auf sich nehmen, ein arbeitsintensives Freifach zu besuchen, betreten damit in zweifacher Hinsicht Neuland: Zum einen müssen sie den Umgang mit dem Rechner bzw. seiner Peripherie erlernen, zum anderen werden sie mit einem ihnen noch unbekannten Methodenkanon konfrontiert <sup>7</sup>. Die Absolventen dieser Lehrveranstaltungen hatten jedoch bisher auf universitärem Boden keine Möglichkeit, nach der Einführung auch weiter ausgebildet zu werden, weder in rein arbeitstechnischer Hinsicht (EDV-Einsatz), noch in methodischer (Quantifizierung). Nur für eine Minderzahl bot sich die Gelegenheit, innerhalb eines Forschungsprojektes von Fachkräften betreut und unterwiesen zu werden und/oder den

---

bis 1250 beschäftigt und am Forschungsinstitut für Historische Grundwissenschaften unter der Leitung von Univ.-Prof. Dr. *Friedrich Hausmann* durchgeführt wird.

<sup>5</sup> Das vorliegende Papier beschäftigt sich mit dem Konzept und ersten Erfahrungen über eine Lehrveranstaltung, die unter der gemeinsamen Leitung von Dr. *Hermine Losch* und des Verfassers im Sommersemester 1985 neben der oben-erwähnten Einführung erstmals abgehalten wird; ihr Titel lautet "EDV und Geschichtswissenschaft II: Von der Quelle zum Datensatz."

<sup>6</sup> Vgl. *Stefan Karner*: Erfahrungen mit Quantifizierung und EDV in der Lehre, In: Bericht über den 16. Öst. Historikertag in Krems / Donau (= Veröff. d. Verbandes Öst. Geschichtsvereine 25), o.O., 1985, S. 684-686.

<sup>7</sup> Zur Situation der Lehre an den österreichischen Universitäten in bezug auf Quantifizierung vgl. *Gerhard Botz*: Quantifizierende Geschichte als Gegenstand (inter-)universitärer Weiterbildung, In: Bericht über den 16. Öst. Historikertag in Krems / Donau (= Veröff. d. Verbandes Öst. Geschichtsvereine 25), o.O., 1985, S. 667-675.

QUANTKURS in Salzburg zu besuchen, dessen Bedeutung für die Quantifizierung in Österreich gar nicht hoch genug veranschlagt werden kann <sup>8</sup>.

Im Anschluß an die Sitzung des Arbeitskreises "Quantifizierung in der österreichischen Geschichtsforschung" auf dem österreichischen Historikertag 1984, in der diverse didaktische Konzepte der Quantifizierung und EDV-Anwendung diskutiert wurden, erfolgte am Ludwig-Boltzmann-Institut für Historische Sozialwissenschaft eine integrative Formulierung jener Grundsätze, die in der Folge auch Basis für die einschlägigen Grazer Lehrveranstaltungen wurden.

## 2. Zielsetzungen

Um die Vermittlung eines einheitlichen Basiswissens über EDV-Anwendung und Quantifizierung zu ermöglichen, sollten universitäre Lehrveranstaltungen demnach folgende Schwerpunkte zum Inhalt haben <sup>9</sup>:

- Einführung in die methodologischen Voraussetzungen
- Quellenkritik
- theorie- und hypothesengeleitetes Entwickeln von Fragestellungen
- Einführung in die (zumindest deskriptive) Statistik
- Einführung in das lokale EDV-System
- Einführung in statistische Programmpakete (SPSS, SPSS-X, BMDP)
- Einführung in ein historisches Datenbanksystem (CLIO)
- Interpretation und Diskussion von Ergebnissen quantifizierenden Arbeitens.

Diese Bereiche entsprechen den Erfahrungen aus dem QUANTKURS und sollen derzeit an den Universitäten Wien, Salzburg und Graz realisiert werden.

Die bereits angesprochene äußerst unbefriedigende Situation an der Grazer Universität im Bereich der EDV-Anwendung und Quantifizierung führte im Sommer 1984 zu Überlegungen, auf welche Weise hier Abhilfe gefunden werden könnte. Aufgrund der obengenannten Schwerpunkte und der Forderung, auch quellennahe und so eher qualitative Datenverarbeitung zu vermitteln, wurde eine Zielvorstellung entwickelt, die in mehreren, aufeinander aufbauenden Lehrveranstaltungen verwirklicht werden soll:

---

<sup>8</sup> Vgl. Gerhard Botz: Quantkurs am Scheideweg, in: Historical Social Research (HSR) 29, 1984, S. 86-95.

<sup>9</sup> Überlegungen zur Integration der Ausbildung in Quantifizierung und EDV-Anwendung für Historiker in den regulären Studiengang österreichischer Universitäten, hrsg.v. Ludwig-Boltzmann-Institut für Historische Sozialwissenschaft, 1984 (unveröffentlichtes Papier).



Vorerst gilt es, grundlegendes Wissen über die EDV als Arbeitstechnik (ob man sie nun als Grund-, Hilfs- oder Nachbarwissenschaft sehen will, spielt dabei keine Rolle) zu vermitteln und deren Anwendungsmöglichkeiten kurz vorzustellen. Damit auf das engste verknüpft ist die Kenntnis der theoretischen, methodischen und konzeptuellen Voraussetzungen quellennaher und methodenorientierter Datenverarbeitung.

Weiters soll der Historiker mit den für ihn spezifischen Programmsystemen umzugehen lernen, nämlich mit dem fachspezifischen und datenbankorientierten Programmsystem CLIO sowie dem Statistikpaket SPSS. Dies soll ihm selbständiges Arbeiten am Computer ermöglichen oder dazu anregen.

Studierende sollen demnach in die Lage versetzt werden, aufgrund ihrer universitären Ausbildung kleinere Projekte von der Seminararbeit bis zur Dissertation selbsttätig zu planen und durchzuführen, und zwar sowohl arbeitstechnisch am Rechner als auch methodisch, z.B. quantifizierend. Daneben sollten sie ohne grössere Probleme kompliziertere Verfahren in beiderlei Hinsicht erarbeiten können, entweder durch Literaturstudium und/oder durch weiterführende Lehrangebote, wie sie etwa die Fortgeschrittenen-Gruppen des QUANTKURSES darstellen <sup>10</sup>.

### **Ausgangssituation**

Das didaktische Konzept einer solcherart vordefinierten Lehrveranstaltung ist natürlich von zahlreichen Imponderabilien abhängig, von denen die Ausstattung und Kapazität des lokalen Rechenzentrums, die Zusammensetzung der Zielgruppe, also der Studierenden, sowie die rechtlichen Grundlagen wie Bewilligung, Anrechenbarkeit im Studienplan etc. herauszustreichen sind. Obwohl hier ein Vergleich mit anderen Universitäten kaum möglich ist, trugen diese lokal bedingten Einflüsse wesentlich zur Konzeption der Lehrveranstaltung bei und bedürfen demnach auch einer Erläuterung. Vor allem sind es jedoch die methodisch/konzeptuellen Vorüberlegungen, die den Rahmen für die Verwirklichung der vorgegebenen Ziele abstecken.

### **Rechenzentrum**

Wie die meisten auf naturwissenschaftliche, also rein numerische Anwendungen spezialisierten Universitätsrechenzentren bietet auch das "EDV-Zentrum der Karl-Franzens-Universität Graz" dem Historiker vorerst nicht

---

<sup>10</sup> Zu weiteren Möglichkeiten, wie etwa Sommerschulen, vgl. *Gerhard Botz*: Quantifizierende Geschichte als Gegenstand (inter-)universitärer Weiterbildung, In: Bericht über den 16. Öst. Historikertag in Krems / Donau (= Veröff. d. Verbandes Öst. Geschichtsvereine 25), o.O., 1985, S. 671.

allzu viel. Die Grazer Universität hat neben anderen Institutionen Anteile an der Rechenkapazität der Großrechenanlage UNIVAC 1100/81 (SPERRY) und besitzt als eigenen Bereichsrechner eine VAX 11/750 (DEC). Beide Systeme sind jedoch durch ihre spezifische Ausstattung an Hardwarekomponenten und Software für die Anliegen eines Historikers keineswegs ideal<sup>11</sup>. Während auf der UNIVAC zwar die Programmsysteme CLIO, SPSS und SPSS-X laufen, erweist sich der zeilenorientierte Editor als äußerst unkomfortabel und das Betriebssystem unterstützt weder den deutschen Zeichensatz noch die für die maschinelle Edition oft unentbehrlichen Sonderzeichen. Zudem ist auch der am EDV-Zentrum installierte Drucker der UNIVAC nicht in der Lage, Groß- und Kleinschreibung wiederzugeben. Im Gegensatz dazu zeichnet sich die VAX durch einen sehr komfortablen Full-Screen-Editor aus, unterstützt seit jüngstem auch den deutschen Zeichensatz, und auf den zur Verfügung stehenden Druckern lassen sich - neben der eigentlich selbstverständlichen Groß- und Kleinschreibung - auch Umlaute und verschiedenste Sonderzeichen darstellen. Leider ist sie ausschließlich mit naturwissenschaftlich orientierter Software ausgestattet, was Anwendung auf nichtnumerischem Gebiet - mit Ausnahme des Editors - derzeit verhindert; allerdings ist die VAX als Remote-Station an die UNIVAC angeschlossen.

Ein weiteres Problem stellen die zugänglichen Kapazitäten der überlasteten und zum Teil nicht mehr ganz neuen Systeme dar. Abgesehen davon, daß die Abarbeitung größerer Jobs sehr lange dauert, sind komplexere Aufgaben innerhalb der bestehenden Limits nur in Einzelfällen zu verwirklichen. Auch der einzige Terminalraum, der mit 15 Terminals (!) ausgestattet ist, verhindert die Abhaltung von Lehrveranstaltungen untertags.

### 3.2. Zielgruppe

Die heterogene Zusammensetzung jener Gruppe von Studenten und Graduierten, welche sich im Sommersemester zusammengefunden hat(te), bestätigte nur die Erwartungen. Der unterschiedliche Ausbildungsstand, nicht nur in bezug auf EDV-Einsatz oder Quantifizierung, sondern auch was etwa paläographische Kenntnisse oder Quellenkritik betrifft, erschwert und verlangsamt das gemeinsame Arbeiten enorm. Was für den Einen absolutes Neuland darstellt, führt beim Anderen zur Langeweile; allerdings bringt dieser Unterschied in den Vorkenntnissen paradoxerweise wieder einen gewissen

---

<sup>11</sup> Über die Novissima auf dem Gebiet der Hard- und Softwarekomponenten berichtet das vom EDV-Zentrum herausgegebene Periodikum "Memory" (hier besonders 6, 1984 und 7, 1985).

Ausgleich: so benötigt der Teilnehmer mit EDV-Erfahrung etwa die Hilfe eines EDV-unerfahrenen Kollegen, um die Quelle richtig lesen zu können.

Auch der Erwartungshorizont der potentiellen Besucher einer solchen Lehrveranstaltung ist schwer abzuschätzen. Während besonders Anfänger eine äußerst diffuse Vorstellung von EDV und Quantifizierung entwickeln, haben jene Kollegen, die in Projekten arbeiten oder den QUANTKURS absolviert haben, sehr konkrete und spezifische Erwartungen. Besonders erfreulich ist die Teilnahme von Studierenden und Graduierten verwandter Fachbereiche wie etwa der Volkskunde.

### 3.3. Studienplan

Noch heute ist die EDV "von einem manchmal lockenden, manchmal abschreckenden, stets aber sehr geheimnisvollen und manchmal geradezu magischen Nebel unvollständiger und ungenauer Information umgeben" <sup>12</sup>. Aus diesem Grund ist es wohl nicht immer möglich, in den Gremien die erhoffte Aufgeschlossenheit zu finden, wenn es um die Einrichtung von einschlägigen Lehrveranstaltungen und deren Anrechenbarkeit geht. Immerhin ist es in Graz gelungen, im Sommersemester 1985 drei solcher Kollegs anzubieten, die als Wahlfach für den Studiengang Geschichte laut Studienplan angerechnet werden <sup>13</sup>. Obwohl von einer Verankerung im Studienplan noch keine Rede sein kann, scheint doch das Angebot an EDV-Lehrgängen für Historiker mittelfristig gesichert.

### 3.4. Methodische/konzeptuelle Vorüberlegungen

Es ist hier nicht der Ort, die Vor- und Nachteile datenbankorientierter Verfahren im allgemeinen oder eines DBMS im speziellen zu diskutieren. Es

---

<sup>12</sup> *Manfred Thaller: Descriptor. Probleme der Entwicklung eines Programmsystems zur computerunterstützten Auswertung von Bildquellen, In: Europäische Sachkultur des Mittelalters, Wien, 1980. (= Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs.4.), S. 167.*

<sup>13</sup> *Karner: EDV und Geschichtswissenschaft I: Einführung in Probleme der Quantifizierung in der Geschichtswissenschaft (mit Einführung in einfache EDV-Programme), UE, 2st.*

*Kropac/Losch: EDV und Geschichtswissenschaft II: Von der Quelle zum Datensatz. VL, 2st.*

*Thaller: EDV und Geschichtswissenschaft III: Quellenkritik und Quellenkunde rechnergestützter historischer Forschung. VL, 2st.*

soll nur kurz skizziert werden, warum diesen Verfahren ein so hoher Stellenwert eingeräumt wird. Wie bereits aus dem Titel der Lehrveranstaltung hervorgeht, soll nicht die Quantifizierung als Methode, sondern der EDV-Einsatz als Arbeitstechnik im Mittelpunkt der Vorlesung stehen. Selbstverständlich muß dabei auch das Hauptanwendungsgebiet, die quantifizierenden Methoden, ausführlich behandelt werden, allerdings sollen auch nützliche Anwendungen ohne Anspruch auf methodische Innovation - wie beispielsweise die maschinelle Verwaltung umfangreicher Bibliographien - Erwähnung finden. Die oft betonte Trennung und ebensooft stillschweigend vorgenommene Gleichsetzung von EDV-Einsatz und Quantifizierung erscheint als problematisch; es ist besonders bei Einführungen wichtig zu demonstrieren, daß eine Arbeitstechnik keine Methode impliziert und umgekehrt. Dieser Lehrveranstaltung liegt daher das ausdrückliche Bekenntnis zur Methodenpluralität zugrunde <sup>14</sup>: der Einsatz des Rechners soll die Anwendung mehrerer Methoden erleichtern oder erst ermöglichen, so beispielsweise die rechnergestützte, aber methodisch traditionelle Hermeneutik neben der Quantifizierung.

Daher sollte in einem ersten Teil die quellenorientierte Datenverarbeitung behandelt werden, was datenbankorientiertes Arbeiten sowohl mit fortlaufenden Texten als auch mit formalisierten Quellen(teilen) bedeutet. Diese Phase soll einerseits die hermeneutische Interpretation unterstützen, andererseits als explorativer Vorschaltprozeß für die Kodierung dienen, um die Quelle besser "kennenzulernen", und das Entwickeln statistisch zu testender Hypothesen zu begünstigen <sup>15</sup>. Die quellenorientierte Datenverarbeitung bildet somit Gegenstück und/oder Vorstufe zur methodenorientierten, mit Hilfe derer die erstellten Hypothesen durch diverse statistische Verfahren getestet werden. Letztere soll demnach in einer zweiten Stufe behandelt werden, besonders, da ein gewisses statistisches Grundwissen unbedingte Voraussetzung ist.

---

<sup>14</sup> Zu diesen Problemen neuerdings Konrad H. Jarausch, Gerhard Armingier, Manfred Thaller: *Quantitative Methoden in der Geschichtswissenschaft. Eine Einführung in die Forschung, Datenverarbeitung und Statistik*, Darmstadt, 1985, bes. S.1-6, 203-206. - Vgl. auch Manfred Thaller: *Zur Formalisierbarkeit hermeneutischen Verstehens in der Historie*, In: *Mentalitäten und Lebensverhältnisse*, Rudolf Vierhaus zum 60. Geburtstag, Göttingen, 1982, S. 439-454.

<sup>15</sup> Vgl. Manfred Thaller: *Datenbankorientierte Verfahren in der Geschichtswissenschaft*, In: Bericht über den 15. österreichischen Historikertag in Salzburg, Salzburg, 1984. (= Veröffentlichungen des Verbandes Österreichischer Geschichtsvereine. 23.), S.433 f.

#### 4. Alternativen zur Verwirklichung

Von J.Ehmer wurden drei grundlegende Alternativen genannt <sup>16</sup>, die bei der Konzeption einschlägiger Lehrveranstaltungen zu berücksichtigen sind:

Fixes Format - freies Format

Kodierte Eingabe - quellennahe Eingabe

Datenfileorientierung - Datenbankorientierung

Diese Alternativen wurden bereits andernorts ausführlich diskutiert <sup>17</sup>. Hier fiel die Entscheidung für eine quellennahe Eingabe im freien Format, die Arbeitsweise soll datenbankorientiert erfolgen. Neben den obengenannten Vorgaben sprachen in der Hauptsache zwei Gründe dafür:

Die Teilnehmer sollen nach Absolvierung der Vorlesung in der Lage sein, beliebige Quellen auch komplexerer Struktur zu formalisieren und maschinenlesbar zu machen, sowie auch mit fortlaufenden Texten arbeiten zu können.

Das derzeit einzige Programmpaket, welches Forderungen dieser Art erfüllen kann, ohne zur Programmierung auf der Ebene Höherer Programmiersprachen zu zwingen, ist das von M.Thaller in Göttingen entwickelte System namens CLIO <sup>18</sup>. Aufgrund diverser Projekte konnte CLIO in Graz implementiert werden und ist dankenswerterweise auch für Dritte benutzbar.

Neben diesen drei Wahlmöglichkeiten ergeben sich bei der Konzeption dieser Vorlesung weitere Optionen, teils aus der speziellen lokalen Situation, teils aus grundlegenden Überlegungen resultierend.

Als erstes stellte sich die Frage, welcher Rechner nun tatsächlich zu benützen sei. Die in Punkt 3.1 angesprochene Verbindung der beiden Grazer

---

<sup>16</sup> Josef Ehmer: Erfahrungen mit EDV-gestützten Projekten als Grundlage für die Konzeption einer Lehrveranstaltung "Einführung in die EDV für Historiker", In: Bericht über den 16. Öst. Historikertag in Krems / Donau (= Veröff. d. Verbandes Öst. Geschichtsvereine 25), o.O., 1985, S. 676-683.

<sup>17</sup> Josef Ehmer: Erfahrungen mit EDV-gestützten Projekten als Grundlage für die Konzeption einer Lehrveranstaltung "Einführung in die EDV für Historiker", In: Bericht über den 16. Öst. Historikertag in Krems / Donau (= Veröff. d. Verbandes Öst. Geschichtsvereine 25), o.O., 1985, S. 680f. - Vgl. auch Donald E. Ginter, Peter Grogono, Frederick A. Bode: A Review of Optimal Input Methods: Fixed Field, Free field, and the Edited Text, in: Historical Methods Newsletter 10, 1977, S. 166-176.

<sup>18</sup> Vgl. Manfred Thaller: CLIO. Einführung und Systemüberblick, Göttingen, 1983 (Preprint). Neben dieser Einführung veröffentlichte M.Thaller weitere, hier nicht angeführte Dokumentationen sowie eine größere Anzahl von Studien zu Konzeption und technischer Realisierung von CLIO.

Rechner macht es möglich, die VAX wegen ihrer moderneren I/O-Komponenten als "Arbeitsplatzrechner" und die UNIVAC als "Hintergrundsystem" zu benutzen. Für den Benutzer ergeben sich daraus nur zwei wesentliche Unterschiede: Batch-Jobs werden statt mit "@START" mit dem Befehl "REMOTE" gestartet, Outputs liegen als Printfile (Filetyp .lis) in der Default-Directory der VAX und können nach Bedarf ausgedruckt werden. Die HELP-Facilities in Betriebssystem und Editor der VAX und der im Vergleich mit der UNIVAC-Steuersprache wesentlich einfachere Aufbau der Digital Command Language kommen den Anfängern besonders entgegen. Die der UNIVAC zu übergebenden Batch-Run-Karten werden nur vorgegeben, die UNIVAC also im Prinzip als "Blackbox" behandelt. Das bedeutet, daß die angeführte Alternative de facto durch die Benützung beider Rechner umgangen wird.

Ähnlich verhält es sich mit der Frage, ob - speziell für statistische Verfahren - ein bestehender Datensatz zu verwenden oder ein neuer anzulegen sei <sup>19</sup>. Das Konzept M. Thallers <sup>20</sup>, bei mehreren QUANTKURSEN bestens bewährt und hier übernommen, behandelt die Errichtung einer SPSS-Systemdatei erst nach der Systemfilebeschreibung und einfacheren statistischen Auswertungen. Dieses Vorgehen bedingt einen bestehenden, bereits getesteten Datensatz, der Variable verschiedener Meßniveaus enthält und die Anwendung verschiedener statistischer Routinen erlaubt; ein solcher stand bei der Planung der Lehrveranstaltung bereits zur Verfügung <sup>21</sup>. Dagegen spricht der Umstand, daß die Studierenden oft langwierigen Weg der Datenaufbereitung bis zum statistisch auswertbaren File kennenlernen und an allen Schritten beteiligt werden sollen. Auch die in Punkt 3.4 angestellten Vorüberlegungen scheinen dem Thallerschen Konzept zu widersprechen. Einen Ausweg aus diesem Dilemma würde die Verwendung derselben Quelle (oder Quellengattung) für die quellenorientierte Datenverarbeitung bringen, wie sie für die Erstellung des bestehenden (statistisch auswertungsbereiten) Datensatzes verwendet wurde. Um sinnlose Doppelarbeit zu vermeiden, könnte u.U. ein anderer Zeitraum gewählt werden, immer unter der Voraussetzung, daß es sich um identische Strukturen handelt. Daraus ergibt sich auch eine Entscheidung in diesem Bereich - es wäre günstig, beide Wege gleichzeitig zu beschrei-

---

<sup>19</sup> Dazu vgl. Thomas Schuler: Quantifizierung und Geschichtsstudium, In: HSR 17, 1981, S. 61f.

<sup>20</sup> Manfred Thaller: Numerische Datenverarbeitung für Historiker. Eine praxisorientierte Einführung in die quantitative Arbeitsmethode und in SPSS, Wien-Köln, 1982 (= Materialien zur Historischen Sozialwissenschaft. 1.).

<sup>21</sup> Herrn Dr. Gerhard Jaritz, Institut für mittelalterliche alienkunde Österreichs, sei für die Weitergabe dieses Datensatzes herzlich gedankt.

ten. Weiters wurde es notwendig zu entscheiden, ob quellenorientiert oder methodenorientiert, ob streng quantifizierend oder methodisch eher neutral vorzugehen sei. Die Beantwortung dieser Fragen ergibt sich zwingend aus Punkt 3.4: es soll sowohl quellen- als auch methodenorientiert und nach mindestens zwei geschichtswissenschaftlichen Methoden gearbeitet werden, ohne einseitige Wertung.

## 5. Inhalt der Lehrveranstaltung

### 5.1. Übersicht

#### *Einführungsphase:*

- *Grundwissen über den Computer (Hard/Software) und dessen Peripherie*
- *Anwendungsmöglichkeiten der EDV in der Geschichtswissenschaft*
- *Probleme datentechnischer Verfahren in der Historie - Theorien, Methoden, Konzepte*

#### *1. inhaltliche Phase - quellenorientierte Datenverarbeitung*

- *Quelle und Quellenkritik; Informationsgehalt und -hierarchie der Quelle*
- *CLIO-Format: Eingabekonventionen, Datenstruktur, Plausibilität und Praktikabilität*
- *Erfassung der Quelle am Bildschirmterminal*
- *CLIO-Systemfilegenerierung (Befehle und deren Elementarmodi); Auswertung des "Prüflaufprotokolls"; wenn notwendig, dann Korrektur der Rohdaten und erneute Systemfilegenerierung*
- *Information Retrieval: Formulierung inhaltlicher Fragen, Übersetzung in die CLIO-Retrieval-Sprachkomponenten;*
- *"Linguistische Datenverarbeitung" im weitesten Sinne: Erstellung von Wortzählungen und eines Wörterbuchs*

#### *2. inhaltliche Phase - methodenorientierte Datenverarbeitung (Quantifizierung)*

- *Einführung in die Quantifizierung (Wiederholung)*
- *Statistisches Grundwissen (z.B. Meßniveau, Stichprobe / Grundgesamtheit etc.)*
- *Vorstellung des Programmpakets SPSS*
- *SPSS-Systemfilebeschreibung*
- *einfache Auswertungen aus dem Gebiet der deskriptiven Statistik*
- *maschinelles Kodieren mit Hilfe der entsprechenden CLIO-Routinen*
- *SPSS-Systemfilegenerierung.*

## 5.2. Praktische Durchführung

Beide inhaltlichen Phasen beschäftigen sich mit derselben Quelle, einem Rechnungsbuch der Chorfrauen von Klosterneuburg, 1495-1508 <sup>22</sup>. Diese Quelle eignet sich durch ihre leicht leserliche Schrift und ihre klare Struktur (Abb. 1) besonders, um dieses Konzept für die eingangs erwähnte heterogen zusammengesetzte Zielgruppe zu realisieren. Die Teilnehmer sollen lernen, den oft langwierigen Weg von der Originalquelle zum (analytischen) Datensatz mit allen seinen Problemen und Tücken selbst zu beschreiten.

In der ersten Phase werden die Jahre 1495/96 des Rechnungsbuchs als Xeroxkopie vorgelegt und quellenkritisch betrachtet. In der Folge wird die Quelle in zweifacher Hinsicht zur Eingabe aufbereitet: einmal formalisiert in einer hierarchischen Struktur für Retrieval-Programme und maschinelle Kodierung, zum anderen als fortlaufender Text gleichsam als Vorbereitung zur Edition, um "Textanalysen" in Form von Wortzählungen, KWICs und KWOCs vornehmen zu können. Bei der hierarchischen, datenbankorientierten Behandlung der Quelle gilt es, deren Informationsgehalt und -struktur logisch und möglichst unverkürzt abzubilden. Das bedeutet eine Differenzierung der einzelnen Informationen nach mehreren Richtungen. Es ist zu prüfen, ob die Information unmittelbar aus der Quelle übernommen werden kann wie hier etwa der Wochentag oder die Angabe des Preises, oder ob sie erst mittelbar aus anderen Angaben zu erschließen ist, wie beispielsweise das Kalenderdatum oder die Fastenregelung. Weiters bedarf es einer Zuweisung der Informationen an die verschiedenen hierarchischen Ebenen, wobei anfangs die Struktur der Quelle möglichst direkt in jene des Datenmodells übernommen werden sollte. Daraus ergibt sich für die vorgestellte Quelle die Struktur von Schema 1 (in der Notation von CLIO).

Die im Quellenbeispiel (Abb. 1) gezeigte Woche würde unter Zugrundelegung des Datenmodells aus Schema 1 aussehen wie Schema 2.

Die Bearbeitung der Quelle als fortlaufender Text erscheint vom Datenmodell her wesentlich einfacher, muß doch lediglich der Text erfaßt werden und der Ort in der Quelle, wo jener zu finden ist (auf die seltenen Textvarianten und alternative Betrachtungsweisen wurde vorerst verzichtet). Wir erhalten dadurch Schema 3.

Dieselbe Woche, die oben bereits formalisiert dargestellt wurde, müßte in der Eingabe dann wie Schema 4 aussehen.

---

<sup>22</sup> Niederösterreichisches Landesarchiv, HS. 424. - Dazu vgl. Gerhard Jaritz: Aderlaß und Schröpfen im Chorfrauenstift Klosterneuburg (1445 -1533), In: Jahrbuch des Stiftes Klosterneuburg N.F. 9, 1975, S.78 f.



D(okument)\$QUELLE/ZITAT

↑

↑↑<< W(ochentag)\$NAME/DATUM/FASTENREGELUNG/

↑↑     ↑ FESTREGELUNG

↑↑     ↑

↑↑     ↑<< E(intragung)\$NUMMER/ANZAHL%EINHEIT/OBJEKT/

↑         RECHTSFORM/PREIS/ZIELGRUPPE/HERKUNFT

↑

↑<< S(umme)\$GESAMTAUSGABEN

Schema 1

Durch einfache Manipulationen läßt sich in der Folge aus einer Dokumentendatenbank eine Wortdatenbank errichten.

Parallel zu Modellbildung und Datenerfassung ist es bereits notwendig, das Wie und Warum der einzelnen Entscheidungen festzuhalten und zu dokumentieren; zum einen muß jeder Mitarbeiter vom anderen wissen, wie er zu seinen Eingabedaten gekommen ist, zum anderen ist eine solche Dokumentation *condicio sine qua non* für eine spätere Verwendung des Datensatzes.

Aufgrund der geschilderten Situation am lokalen Rechenzentrum sind die Studierenden angehalten, nach einem Einführungsabend an den Terminals alle Manipulationen in Zusammenhang mit der Datenerfassung selbständig durchzuführen. Lediglich die Kollationierung der Quelle und damit verbunden die Vorgabe der Korrekturen werden von den Leitern der Lehrveranstaltung vorgenommen. Dieser Zwang zur selbständigen Arbeit erfordert natürlich eine ständige begleitende Betreuung der Teilnehmer, die ihre Probleme hauptsächlich in den Lehrveranstaltungsblöcken, aber auch außerhalb der regulären Vorlesungszeiten mit den Vortragenden und untereinander besprechen sollen. Auf diese Weise wird der Notwendigkeit entsprochen, möglichst wenig Gruppenarbeit am Rechenzentrum zu absolvieren. Die Erarbeitung der Aufgaben und Programme, sowie deren Analyse und Interpretation erfolgt ausschließlich am Institut; lediglich drei Arbeitsabende am Rechenzentrum sollten ausreichen, den Umgang mit Hardware und Software zu erklären bzw. zu vertiefen. Das bedeutet zwar die Trennung von theoretischer und praktischer Arbeit, bereitet aber andererseits auf den Alltag der EDV-Anwender vor, auftretende Probleme ohne unmittelbare Beratung aus eigener Kraft lösen zu müssen. Um die Eingabearbeit und in der Folge auch

D\$RECHNUNGSBUCH KLOSTERNEUBURG 1495-96/NÖLA HS. 424 FOL. 18'  
 W\$SONNTAG%p(ri)ma po(st) Trini(ta)te/1496.6.5/N/F  
 E\$1/14/HÜHNER/K/3S. 6D.  
 E\$2/1/SCHWEINEBRATEN/K/70D.  
 E\$3//SEMMELN/K/23D./GÄSTE  
 E\$4/3/GEKRÖSE/K/12D.  
 E\$5/0.5/KALB/K/3S. 20D.  
 W\$MONTAG/1496.6.6/N/W  
 E\$1//SCHWEINEFLEISCH/K/32D.  
 E\$2/2/EUTER/K/3S. 20D.  
 E\$3/2/SPANFERKEL/K/32D.  
 E\$4/2/HÜHNER/K/15D./KRANKE  
 W\$DIENSTAG/1496.6.7/N/W  
 E\$1//SELCHFLEISCH/E  
 E\$2/0.5/SCHAFBAUCH/K/36D.  
 E\$3/1%VIERTEL/KALBFLEISCH/K/60D.  
 W\$MITTWOCH/1496.6.8/N/W  
 E\$1//MILCH/K/12D.  
 E\$2/1%ACHTTEL/MILCHRAHM/K/32D.  
 E\$3//KREBSE/K/12D.  
 E\$4//EIER/K/53D.  
 W\$DONNERSTAG/1496.6.9/N/W  
 E\$1/0.5/KA STRAUN/K/38D.  
 E\$2/7/HÜHNER/G  
 E\$3/1/KALBSBRUST/K/66D. !  
 E\$3/1%VIERTEL/KALBFLEISCH/K/66D. !  
 W\$FREITAG/1496.6.10/N/W  
 E\$1//KREBSE/K/68D.  
 E\$2//MILCH/K/12D.  
 E\$3//ZWIEBEL/K/8D.  
 W\$SAMSTAG/1496.6.11/N/W  
 E\$1//EIER/K/77D.  
 E\$2//MILCHRAHM/K/46D.  
 S\$4TL. 36D.

Schema 2

D(okument)\$FOLIERUNG

↑

↑<<\$ZEILE/TEXT

Schema 3

die Erstellung von Befehls- und Runkarten zu erleichtern, arbeiten jeweils zwei Studierende an einer Aufgabenstellung.

In der zweiten Phase wird das Jahr 1497 des obengenannten Rechnungsbuchs bearbeitet, das als fertiger - am QUANTKURS getesteter und bestens bewährter - Datensatz vorliegt. Da der Inhalt der Quelle innerhalb der beiden Jahre kaum differiert, kann durch die vorausgehende Bearbeitung der Jahre 1495/96 in der ersten Phase ein verbessertes inhaltliches Verständnis jener statistischen Verfahren, denen der Datensatz unterworfen wird, sowie der folgenden Kodierungsarbeit erreicht werden. Die maschinelle Kodierung der eingegebenen Daten, die Übersetzung der Codes und die Errichtung eines SPSS-Systemfiles sollen den Grundkurs abschließen.

## **6. Zukunftsperspektiven**

### **6.1. Ein Grazer Curriculum?**

Wie eben durch die Bezeichnung Grundkurs angeklungen ist, soll auf Wunsch der Studierenden eine Fortsetzung angeboten werden, die allerdings organisatorisch und konzeptuell wesentliche Abweichungen vom bisher erörterten Modell enthalten muß. Eine weiterführende Lehrveranstaltung sollte verstärkt inhaltlich orientiert sein, d.h. vermehrt auf das Entwickeln von Hypothesen und Fragestellungen abzielen und die im Grundkurs erlernten Routinen anhand inhaltlicher Schwerpunkte vertiefen. Weiters sollte es den Studierenden ermöglicht werden, eigene Quellen, Fragestellungen und Probleme einzubringen und zu diskutieren. Die Organisation als Blockvorlesung soll auch in der Fortsetzung aufrecht bleiben.

Die in Punkt 3.3 vorgestellte Abfolge der Lehrveranstaltungen soll in der nächsten Zeit beibehalten werden. Stufe I ermöglicht ein erstes Hineintasten in die EDV und Quantifizierung, ist für Anfänger konzipiert und wird jedes Semester gelesen. Die hier diskutierte Stufe II soll sich über zwei Semester erstrecken und bietet zwar auch die Einstiegsmöglichkeit für Anfänger, geht aber wesentlich intensiver vor und benötigt einen nicht geringen Arbeitsaufwand. Stufe III vermittelt als reine Vorlesung theoretisches Wissen auf hohem Niveau; je nach Verfügbarkeit geeigneter Lektoren sollte diese Vorlesung doch

D\$18'

\$1/Do(m)i(ni)ca p(ri)ma po(st) Trini(ta)te

\$2/Item umb 14 huendel 3 ß 6 d

\$3/Item umb 1 sweinenpraten 70 d

\$4/Item umb seml den gesten 23 d

\$5/Item umb 3 kröß 12 d

\$6/Item umb 0.5 kalb 3 ß 20 d

\$7/Mantag

\$8/Item umb sweinesfleisch 32 d

\$9/Item umb 2 au(o)tter 40 d

\$10/Item umb 2 spensaw 60 d

\$11/Item umb 2 huendel den kranken 15 d

\$12/Eritag

\$13/Item geselichtsfleisch hab ich gehabt

\$14/Item umb 0.5 schofpauch 36 d

\$15/Item umb 1 viertl kalbfleisch 60 d

\$16/Mitichen

\$17/Item umb milich 12 d

\$18/Item umb 1 achtel milichra(e)n 32 d

\$19/Item umb kreussn 48 d

\$20/Item umb a:yr 53 d

\$21/Phintztag

\$22/Item umb 1 kastraun 38 d

\$23/Item 7 huendel hat man uns geschenkt

\$24/Item umb 1 kelbreineprust und ain viertel kalbfleisch 66 d

\$25/Freytag

\$26/Item umb kreussn 68 d

\$27/Item umb milich 12 d

\$28/Item umb zwival 8 d

\$29/Sambstag

\$30/Item umb a:yr 77 d

\$31/Item umb milichräm 46 d

\$32/Summa facit 4 t 36 d

#### Schema 4

zumindest jedes zweite oder dritte Semester mit wechselnden Schwerpunkten abgehalten werden. Ob das derzeitige Angebot den Anforderungen, Historiker in "neuen" Methoden und Arbeitstechniken grundlegend auszubilden,

gerecht werden und sich dadurch auch in Graz etablieren kann, wird erst die weitere Zukunft zeigen. Mit einer Integration in den Studienplan kann kurzfristig jedoch kaum gerechnet werden; die Ausbildung in dieser Richtung beruht auf Freiwilligkeit.

## 6.2. Lehrpaket

In der geschilderten Weise soll sukzessive das gesamte Rechnungsbuch aufgearbeitet und als überschaubare "Mini-Datenbank" für den Lehrbetrieb adaptiert werden. In ein solcherart erstelltes Lehrpaket müßte neben den zur Diskussion gestellten Intentionen der Vortragenden aber auch die Rezeption der Inhalte durch die Studierenden eingehen. Dies soll die Protokollierung jedes Vorlesungsblockes durch dessen Teilnehmer und intensive Diskussion innerhalb der Gruppe gewährleisten. Die auf dem Rechner erstellten Protokolle werden nach Abschluß des Kurses von den Vortragenden redigiert und sollen zusammen mit dem Vorbereitungsmanuskript unter Einarbeitung der wichtigsten einführenden Literatur ein Skriptum bilden, das den Studenten in der Folge zur Verfügung gestellt werden soll. Dieses Skriptum mit der edierten Quelle im Anhang muß auch Beispiele für Befehlskarten und Outputs der speziellen Programme enthalten. Daraus könnte sich nach weiterer Diskussion und den sich daran knüpfenden Umstellungen ein Lehrbuch entwickeln, das seinerseits wieder die Basis eines Lehrpakets bildet. Ein vollständiges Lehrpaket müßte demnach folgende Komponenten enthalten:

1. Lehrbuch mit eingearbeiteten Beispielen und Literaturangaben; edierte Originalquelle.
2. Kommentar zum Lehrbuch für Vortragende, in dem das grundlegende Konzept transparent und ev. Vorschläge für die Realisation gemacht werden,
3. Magnetband mit Rohdaten (Datensatz), Systemdateien (Datenbanken), Befehlskarten und Outputs analog zu den im Lehrbuch enthaltenen Beispielen, sowie der maschinell edierten Quelle.

Die Konzipierung und Einrichtung eines Lehrpakets für andere Großrechner als die UNIVAC 1100/xx oder Mikrocomputer wird vor allem mit den zukünftigen Strategien von CLIO abhängen; eine portable Version des Programmsystems ist in Entwicklung.

Abschließend sei angemerkt, daß die unterbreiteten Vorschläge keinesfalls für sich in Anspruch nehmen wollen, endgültige Modelle und programmatische Zielsetzungen zu präsentieren. Eine Möglichkeit von sicherlich vielen soll hier zur Diskussion gestellt werden, auf welche Weise Benützung und

Aufbau von Datenbanken die universitäre Lehre, aber auch die Forschung erleichtern und fördern könnten. Alle Überlegungen zur Lehrveranstaltung selbst gelten natürlich auch für ein eventuell daraus resultierendes Lehrpaket; Sinnhaftigkeit und Verbesserungsmöglichkeiten am Konzept und dessen Realisation gilt es in der Folge zu prüfen.

Von pma py temit

11	vom xij hündel	xij 3
11	vom Dreierpfeil	lx 3
11	vom Hund den gesen	xxij 3
11	vom ij Feß	lx 3
11	vom j halt	xij 3

Mantag

11	vom Dreierpfeil	xxij 3
11	vom ij Mittel	xl 3
11	vom ij Spensel	lx 3
11	vom ij hündel den kants	xv 3

Erntag

11	Gefelstapfel	lx 3
11	vom j Schopman	xxv 3
11	vom j vielte Ruffel	lx 3

Orntag

11	vom Mittel	xij 3
11	vom j Ochse Mittel	xxv 3
11	vom Ruffel	xl 3
11	vom Ruffel	lx 3

Princktag

11	vom j Ruffel	xxv 3
11	vom j hündel den man	xxv 3
11	vom j keltreierpfeil	lx 3
11	vom j Ruffel	lx 3

Freitag

11	vom Ruffel	lx 3
11	vom Mittel	xij 3
11	vom Ruffel	lx 3

Sonntag

11	vom Ruffel	xxv 3
11	vom Mittel	xl 3

Omne fat mit xxv 3

Abbildung 1

DON KARL ROWNEY

## **The Microcomputer in Historical Research: Accessing Commercial Databases**

The objective of this essay is to describe the uses to which large automated information files may be put by historians who have access to microcomputers. The information files to which I am referring are generally bibliographic files that contain many thousands of citations relevant to the research interests of historians. Some of them exist in both printed and computerized, or "on-line" versions. However, since I want to discuss the use of these files in the context of such criteria as the speed and efficiency with which they can be used, I will discuss only the on-line version of any given file or information service.

### **Access to Large Computers through Desktop Terminals**

The practice of using an electronic computer terminal is familiar to anyone who has used an automatic bank teller or a "Touchtone" telephone. Moreover, by now most people have seen the personal desktop terminals that look rather like a typewriter with a television screen or a printing device and that are attached to a large computer. Such devices have been around college campuses and in such places as banks, police departments and military installations since the 1960s.

Even in the sixties computers were much faster at doing certain things than the human beings who used them or than many of the machines (such as printers and information storage devices) that were attached to them. It was obvious that, under proper control (programs and hardware), substantial increases in efficiency could be realized in some applications by making it possible for several human users to work with one machine simultaneously.

Of course, given the von Neuman "architecture" or internal structural characteristics of computers up until very recently, the experience of simultaneous use by several persons was really an illusion. In fact, the computer could only handle one task at a time; the illusion was produced because the tasks as defined by the users could be subdivided by the computer into a large number of segments each of which could be performed very quickly - quickly enough that blocks of small tasks could be performed for each of several users with no noticeable delay for any of them, and often in a highly complex series of alternative orders.

One could request access to, or loading and analysis of, many different bodies of data at the same time. On the same computer at apparently the



same time, a university budget could be processed, a problem in astronomy (if it wasn't too complex and so didn't require too much memory) could be run, student exams could be graded, and the social structure of, say, Salem, Massachusetts could be analysed. That is, each of several users could enter the relevant data to these problems together with instructions for their analysis at the same time. Alternatively, they could have entered the data (say, the Salem demographics) earlier, and then simply "call" them up from some storage location just as the university treasurer could call up previously entered budget figures for analysis.

The metaphor that comes to mind to illustrate this operation is not that of the one-man band, since the tasks are performed only apparently simultaneously. Rather one might imagine Charlie Chaplin in *Modern Times* rushing from position to position in the factory, performing first one little task and then another. Unlike Chaplin, however, the computer would not grow tired, fall behind, or give up in despair.

## Timesharing and On-Line Computing

In the early 1960s, IBM introduced its multiple user or timesharing system. This system supported (or recognized and operated with) certain groups of typewriter-style terminals (hardware) and instruction languages such as FORTRAN and BASIC (software). Thus, a large computer located, say, in Schenectady, NY, could maintain large bodies of data useful to many different users in many different locations. By using certain commands in FORTRAN or BASIC these users could all access the same database or different databases.

An important feature of such systems, moreover, was that the communication link between the many individual users and the big computer could be a "common carrier." That is, it was not necessary for the user's terminal to be wired directly to the computer. With additional program instructions and a small amount of additional hardware, the terminal, say, in San Francisco or Toledo could use the big computer and its database in Schenectady by sending and receiving the information over a simple telephone line.

More than twenty years ago, the experiences outlined above were re-organized into a modified approach to simultaneous access by several users. It was recognized that individual, large organizations – say a research-oriented think tank or the Internal Revenue Service – made use of many different types of data but that, often, these data had much in common broadly speaking in spite of their differences in detail. Thus, it seemed reasonable that, if these different bodies of data could be formatted for use in the computer

in a way that was highly similar from one body to the next, it ought to be possible to integrate them all into a single data system common throughout the organization.

As a concept, this idea of common formatting is quite simple. As a specialist in Soviet history, I maintain several research files. Some are organized by the names of individuals - Sukharin, Trotsky, Dzerzhinskii, for example. Others are organized by subject - industrial administration, military organization, political parties. The reason why these files are organized so differently is because I assembled them at different times for somewhat different purposes. It is obvious that, so long as these files are maintained separately I will have to go through each of them in order to retrieve all of the information I have on, say, the competition between Trotsky and Bukharin in the industrialization debate of the mid-1920s. Even then, I may miss important pieces of information, since I can't be sure that I will find everything in the individual name file that is relevant to the topic of industrialization any more than I will be certain that I'll find everything in the subject file that is relevant to, say, Trotsky.

On the other hand, there is a way to sidestep the indexing problem. If all my files were organized in roughly the same general fashion (e.g. with authors in one place, titles in another) and if each had a short, descriptive paragraph attached, I might then develop a system of scanning the files, quickly looking for certain terms that were relevant to my specific research interest. When I came upon, or "hit" a card that contained those terms either in the title, say, or in the descriptive paragraph, I could pull it out for further examination and for comparison with other hits.

The problem with the procedure I have just described, of course, is that it will be time consuming if I do it carefully or highly uncertain if I do my scanning very rapidly. This is true even though I may speed things up a bit by placing each different type of information in the same location on all of my cards. This, of course, is the reason why all scholars try to guess what their research interests will be before they structure their data files: in the interests of search efficiency, they rely on indices.

On the other hand, for a computer, "reading" machine-readable data files, functionally indistinguishable from my written card files, is a comparatively fast process. Even in the early days of time-sharing on big computers and of on-line, multiple user databases, it was recognized that data files which were massive in the Department of Defense could be formatted so that they could be reduced to a few large files. It was also recognized that these files could be "read" by a computer that was looking for certain specific pieces of

information as identified by a given user whether or not that user's research interests were reflected in the index terms under which the file was organized.

For example, let us suppose that I was interested in the Korean War and that over five years I had amassed a research file consisting of 10,000 notes I had made during my library research and interviews. If I wanted to check that file for any information that might be relevant to the conflict between President Truman and his commanding general in the field, MacArthur, I would either have to have indexed the topic Truman-MacArthur Conflict when I created the file or I would have to go through each of my 10,000 notes trying to find cards that mention some aspect of the conflict - say Truman and MacArthur in the same context, civilian control over the military, Pentagon and White House, and so forth. In the one instance I would be using a pre-determined analysis structure and I would have to know in advance, what topics and combinations of topics I am likely to be interested in. In the second case I would only have to know that the likelihood is great that any note of relevance to my topic would be likely to mention certain aspects of the problem in the title of the note, other identifying material about the note, or in the body of the note itself. In the first instance I would have established a pre-defined index of the topics that I predict I will be interested in and I would have catalogued my file in terms of that index. In the second instance I would literally scan through the entire file looking for specific terms or combinations of terms (dates, names, events) that would lead me to believe that the note was relevant to my research problem. Most multiple user, on-line, interactive computer systems make use of both of these searching techniques.

Unified formatting and standardization of commands for controlling data stored in computers makes the kind of text searching that I have just described feasible because I could inform the computer, using a standard command language, that I was interested in a specific topic - Truman MacArthur conflict - and it could search each segment of the file items - because of the standard formatting - very rapidly. Obviously this can be done not only for files containing 10,000 items but for those containing hundreds of thousands and more. Indeed, the larger the file the more useful will be the computer.

Moreover, a user who has not actually created the file but who is familiar with the computer searching techniques and with the general pattern according to which the data are tagged or described can call the file up on his or her terminal and use it as though it were her/his own. What makes this sort of event possible is the technology upon which multiple user computer systems rest and the technology upon which uniformly formatted databases rest.

The existence of apparently simultaneous multiple user systems, moreover, has meant that different segments of large standardized databases can be called up by many different users simultaneously and put to several different uses simultaneously. Some of the earliest applications of these concepts were to be found in data-rich, computer-rich (and, of course, cash-rich) environments such as the Pentagon and defense-oriented organizations such as Lockheed Aerospace, Inc. It is thus not accidental that one of the earliest and best-known examples of large, standardized, computer accessible databases is DIALOG, a service of DIALOG Information Systems which is a subsidiary of Lockheed Aerospace.

### **Databases and the Microcomputer. Some Examples**

The advent of the microcomputer has added a new dimension to large, standardized databases. At the present time this is mainly because people who buy such machines for other purposes (e.g. word processing or game playing at home) find that with a little modification the same machine can also be used as a terminal to gain access to large, remote databases and computers. I will describe three of the most common applications of this sort, used by those who have access to terminals or modified microcomputers and telephone links to the appropriate facilities. Keep in mind that, with proper modification, terminals and microcomputers are interchangeable for these purposes.

First, there are the standard and long-established in-house (i.e. proprietary or non-commercial) systems. Most commonly, for academic researchers this application is designed to facilitate access to data that the user himself has stored and to other facilities such as sophisticated statistical packages that are available only via a large computer. Members of university communities often use their institution's computers working from terminals in their offices or even their homes.

Second, there are menu-driven commercial databases accessible via teletext ("Teletex") or similar highly simplified conventional systems. These databases make blocks of comparatively well-defined information available to a user for a more-or-less fixed hourly fee depending upon the time of day that the service is used. One of the largest menu-driven systems in the United States is CompuServe, Inc., a service that operates out of Columbus, Ohio.

Generally, the databases in menu-driven systems run the gamut from aviation weather reports to Dow-Jones financial reports to daily summaries of the news appearing in national newspapers. They also include "network" features such as C-mail or computerized mail services in which one user is

able to send a message addressed to another user who is known to use the same database service. When the addressee enters into the system a symbol or message informs him that "mail" is waiting. Other features include shared computer programs, shared games, and billboards that are similar to mail but that scroll across the screen for any user to see.

As suggested above, different segments of such systems are accessible via a "menu", i.e. a short list of standard commands that will call up a specific data base or transfer one to additional menus.

Finally, and, for our purposes, most important there are research-oriented commercial or public database systems such as Mead Data Central, BRS (Bibliographic Retrieval Services), MEDLARS (National Library of Medicine), and DIALOG. In a reflection of their origins, these databases are oriented toward business, think-tank, government, and, recently, library applications rather more than toward individual users although about one year ago DIALOG introduced a non-primetime service called "Knowledge Index" that is attempting to attract individuals.

These research oriented databases are enormous. They include dozens or scores of files in literally hundreds of fields and thousands upon thousands of individual items. Their potential as sources of information to historians, humanists, and social scientists working independently and in universities is astounding.

The reasons why this is so may be summarized rather simply. First, the information equivalent of a huge - truly exceptional - reference library is available to any scholar regardless of where he or she is, provided only that there are a telephone and an appropriately equipped microcomputer available.

Second, the nature of the on-line system is such that searching is both extremely fast and thorough by comparison with searching many printed sources. In sharp contrast to searches of printed sources, one runs out of ideas, strategies, or paradigms for searching long before the fatigue and ennui associated with plowing through indexes sets in. As a consequence, searches that formerly would consume one or more days can now be completed, literally, in minutes.

Third, searches conducted in conjunction with appropriately equipped microcomputers have the enormous advantage of automatically saving the content of the search. This means that the most time-consuming and laborious part of the bibliographic research process - taking careful notes - now occurs at the rate of thirty or more characters a second without the intervention of the researcher. These stored notes or records can be entered into a permanent data file on the computer, printed for direct entry into a

bibliography, or treated in one of several other ways. Recently, for example, I needed to assemble a bibliographic file consisting of the publications of a dozen active Soviet historians. I assembled the necessary information in my own study in less than two hours by instructing the host DIALOG computer to search several thousand records in the Library of Congress card catalogues (LCMARC, and REMARC), *Historical Abstracts*, *Books in Print*, and the *Social Science Citations Index*.

The total cost of my research on the DIALOG system was about \$130, a sum that was covered by a budget item in a grant under which I was working. To most scholars, used to "free" library reference rooms, \$130 seems a great deal of money for only two hours of computer time. However, anyone who has made extensive use of one of these systems will recognize that two hours is an enormously long time for computerized searching. In my case, moreover, the alternative was to spend two to four days searching the thousands of citations in the relevant printed sources myself—time for which I would not be compensated and which was not covered by my grant.

In any event the process of search is summarized by the term "search strategy" which combines the application of a command language, the choice of one or more databases, and a decision about what to do with the resulting information — i.e. whether simply to look at it on the screen, to store it on a disk, to print it, etc., and, finally, some determination whether the search is complete or whether important material has been omitted through faulty searching or for some other reason.

### Using a System such as DIALOG

The reasons for the institutional orientation of systems such as DIALOG are multiple. For one thing there is the fact that the apparent cost is prohibitive. Databases are priced individually by the hour and sometimes by the accessed record. Prices range from as little as \$35/hour for a file such as the United States Government Printing monthly catalogue (DIALOG file no. 66) to \$300/hour for a file that gives access to chemical and chemically related patents (CLAIMS/UNITERMS DIALOG file 223-225). Individual printed records can be free in certain cases or they can cost as much as \$50 each. Thus, to the inexperienced and timid, the threat of a really big access bill is genuine enough to be disuasive.

Second there is the problem of equipment. Although DIALOG can handle most of the modified microcomputers that a more consumer oriented system such as COMPUSERVE can handle, that fact is not apparent to the user through most of DIALOG's promotional literature. Basically,

DIALOG and similar systems have not seemed to be particularly interested in attracting individuals who are equipped with microcomputers instead of video or printing terminals specifically designed for communication with big computers.

Finally there are problems of access narrowly defined. I mean the problem, for example, of deciding which database to use and then of making efficient use of it so that the twenty or thirty records that one is looking for can be economically extracted out of the 80 million records that are stored. In this case one of the most formidable problems consists in learning a command language. This is a simple language that has been developed to give the host computer instructions about choosing a database and instructions on what to look for once it has selected a database. DIALOG's command language consists in something on the order of 100 individual commands. A related problem comes from applying the command terms to a search so as not to waste time (time is literally money here) and to produce to required information - nothing extraneous, but everything that is available in the databases.

Basically, the search procedure requires the same kind of patience that those who use a micro-computer for word processing need to have in order to make use of the word processing program. The similarities extend to the fact that word processors often combine a menu system with a fairly simple command language. The one I am using as I write this essay, for example, has about sixty individual commands and a very small number of menu items that appear at the bottom of my monitor screen.

Importantly, the orientation process requires not only a knowledge of the command language and of the available databases but also of the research context (e.g. the historical context) of the items being searched. Professional knowledge of the subject being researched substantially improves one's chances of choosing databases correctly and of creating appropriate search strategies within those databases. I have personally witnessed many dollars worth of database time being wasted by individuals who know the appropriate command language and other technical features quite well but who are insufficiently familiar with the context of their problem.

To cite a simplistic example, a searcher who has been asked to assemble a bibliography on the origins of the Russian Revolution of 1905 may not realize that searching *Historical Abstracts* under "REVOLUTION (W) RUSSIA?" will not yield satisfactory results. This is because the vast majority of the "hits" or items that have been selected will deal with the Revolution of 1917. Anyone could make this mistake initially, but a trained scholar would immediately recognize the inadequacy of the search profile and would re-

define it. My experience has been that those who are not trained as scholars in the field of research are considerably less capable of making such a judgement in a timely manner.

It is for this reason - ultimately an economic reason since time and money are wasted - that I argue that in the long run the services that are now being installed by university libraries whereby searches are performed by librarians at the instructions of researchers will be replaced by direct access systems in which the scholars themselves do the searching.

My comments about the forbidding dimensions of a system such as DIALOG for individual or small-scale users are not meant to imply that such systems are inaccessible to individual users who have microcomputers. Quite the contrary. In fact, I have been using DIALOG as an individual (i.e. independently of my university or any other organization) since I first realized that by instructing others to do my searching for me I was only scratching the surface of a huge, important body of information. Apart from the several advantages listed above that have to do with speed, comprehensiveness, and convenience, there are specific advantages for an individual who uses the DIALOG system in the form of (almost) 24-hour accessibility, the fact that there is no minimum monthly charge, the fact that the system is self-documenting for those who are just slightly adventurous (i.e. no training is required - I haven't had any, for example) and that with proper modification, a home or personal computer works fine with the system.

All that is necessary to begin is an application to DIALOG Information Systems, an ability to establish credit, the patience to read through the generally easy to understand documentation that is provided by DIALOG and sufficiently strong nerves to overcome the fear that, in the first two or three sessions, you will run up a bill of, say, \$2,000. Alternatively, if one cannot overcome the fear, or if one is not a good self-instructor, DIALOG provides training sessions (for a fee) all around the United States, in Europe, Canada, Latin America, and Australia several times each year. In addition it provides users with extensive printed documentation about the system as a whole and about each individual data base. Eventually, it seems to me, it should be the role of university libraries to instruct individuals in the use of these systems in much the way that the Library of Congress reference librarians now instruct individuals who use the LC on-line card catalogues.

### **Using the Microcomputer as an On-Line Terminal**

It is clear that a system such as DIALOG was not developed (as was COMPUSERVE) with the microcomputer in mind. On the contrary, the idea was evidently that access would be achieved through one of many dif-



ferent kinds of video or print computer terminals such as those produced by Digital Equipment Corporation, IBM, or Honeywell. Nevertheless, the microcomputer can be used with great effect as a conversational terminal and, by comparison with many terminals, the micro possesses certain distinct advantages since it possesses a memory and, often, a printer that will reorganize and store files in different ways.

First, let me explain the process of transformation. As one will have gathered, the essentials in conversing with a big computerized database are a communications line (e.g. telephone line), and a combination of software and hardware at each end that will translate signals from the terminal/computer into signals transferrable across telephone lines. This latter implies, of course, that translation (or encoding) will work in both directions and both for transmission of data and their reception. In other words, it has to be possible for signals generated by the terminal to be translated into transmissible signals, and for the transmitted signals to be re-translated into signals acceptable to the receiving computer. Similarly, the whole process must also work in reverse.

An important point here is that, because of the transformation of signals that occurs at each end, it is not essential that the devices at both ends (i.e. terminal or computer) speak identical languages or generate identical signals. What is vital is that each must be equipped with a device that will translate its signals into the common signals transmissible across telephone lines and also with a device that will translate telephone signals into signals understandable by it. The devices most commonly employed for this purpose (in the cases of both terminals and microcomputers) are a circuit that works in the terminal itself called an RS-232 serial port. This bit of electronics actually transmits signals out of the terminal, microcomputer, or large computer. However, since telephone lines are designed to transmit only certain categories of signals, a second device is necessary. It most commonly consists of a combination of circuitry and acoustic equipment that transforms the signals generated by the terminal or computer into a form that is transmissible across the telephone lines. This device is called a MODulator-DEMODulator or a MODEM.

A final and highly important ingredient is one or more bodies of instructions that control the incoming and outgoing signals and that may do a number of other things in addition. In the case of microcomputers, these instructions usually take the form of software (disks or cartridges) that must be entered into the microcomputer before initiating communication with the host computer.

To summarize, then, thinking now of the microcomputer as a commu-

nications device, it is clear that several elements must be combined in order to gain an adequate level of terminal-like operation. Obviously, the user will have to be able to use the keyboard and interpret the screen in a fashion appropriate to applying the software and hardware. Also, the software must be adequate to the task of instructing the computer to handle itself in a fashion that will be familiar both to the user and to the host computer. Next, the transmission equipment must be in place (RS-232 and MODEM) and a telephone line (sometimes long-distance) must be available. At the same time a roughly parallel concatenation of devices and instructions must be in place at the opposite end of the communication process.

Assembling the necessary software and hardware is much more difficult to describe in writing than it is to do. I believe that every microcomputer that has been manufactured in substantial amounts has such equipment and software available. I am including here not only the relatively "professional" devices such as Compaq, DEC, IBM-PC, and Apple, but also the distinctly home computers such as Commodore, TI-99, and others. Properly equipped these machines will certainly access menu-driven databases such as COMPUSERVE and they will probably all access the large broadgauged research databases such as DIALOG.

On the other hand a microcomputer equipped for communication may - or may not - be able to function as terminal to a computer system that has its own locally designed terminal system. Such systems would include university systems and systems designed for individual companies, for example. Whether a given configuration of microcomputer, communications hardware and software will work or not will depend on the characteristics of the local system and, especially, on the characteristics of the microcomputer and the software that governs its terminal emulation. For example, I use an Apple II+ as a terminal for all of my database communications functions including access to my university's large computers thirty miles away from my study. However, the simple and very easy to use software that arrived with my "AppleCat II" communications system works splendidly with DIALOG and COMPUSERVE but not with my university's DEC and IBM systems although these are terminal-oriented systems. In order to create a link with those systems I use a much more sophisticated package of communications software (not hardware) that turns the Apple into a far more complex terminal than is demanded by the big commercial databases. Unfortunately this software is considerably more complex to use than the package that arrived with the AppleCat.

## **What are the Important Differences among Representative Databases**

Although my original objective was to discuss "accessing" – not "assessing" – commercial databases, I do want to discuss briefly the differences of some of the big, individual databases as research tools.

Let me begin by observing that my very first experience with on-line research in a large databases was in Washington, several years ago. In the card catalogue room of the Library of Congress I made my first acquaintance with such a system via MUMS. This system is relatively easy to use and highly rewarding for those who do so. It has the important advantage of being free to the user and the terminal area is staffed by a small group of knowledgeable and usually helpful reference librarians. These two characteristics establish a standard for reference assistance by which every library should be judged. As a tool for research in recent bibliography, I believe the LC file is unexcelled, and I speak on the basis of experience with many major libraries here and abroad. In fact the advantages of the LC system are similar to those of other good, research-oriented on-line databases.

While the Library of Congress system is the electronic equivalent of a library card catalogue, the electronic files listed below tend to be more subject specific. All of the following files are accessible via DIALOG. I will comment briefly on each. The reader should keep in mind that the charges listed here are hourly rates but that, for a trained and prudent scholar, it is a rare search that lasts more than a few minutes.

- **Social SCISEARCH (Social Sciences Citation Index DIALOG # 7; \$110/hour).**

Based on the notion that the bibliography that is relevant to a given topic will be found in the titles of the published research and the research citations on the topic. Thus one can begin with a piece of research and trace its intellectual provenance by identifying the works that are cited in its notes, the works that are cited in the notes of those pieces, and so forth. Alternatively, one can search time periods and subjects through titles and subtitles. One can also build up a bibliography for a given author by searching under his/her name. Using this file can be tricky and expensive but also uniquely rewarding. I recommend experience with the print version before using the automated version.

- **Historical Abstracts (DIALOG # 39, \$65/hour).**

Since it is based on searches of author, title, an index and abstracts HA is relatively easy to use and very close to one's previous experiences with author/title and subject headings in card catalogues. Abstracts can be

extremely useful. Considering the straightforward use techniques, it can be relatively cheap and it is certainly a good file to experiment with. One can also use HA on-line in combination with the printed version, since the automated file can identify the sought after items quickly and one can then examine the tagged items at leisure in print. A final, enormous advantage of both the printed and on-line versions of HA is that it is international and multilingual in scope. A major disadvantage of the on-line version of Historical Abstracts is that only comparatively recent files are maintained.

- **Dissertation Abstracts Online (DIALOG # 35; \$72/hour).**  
Rather straightforward to search and relatively inexpensive. This file includes dissertation citations going back to 1861 but abstracts are available only for dissertations entered after January, 1980.
- **Book Review Index (DIALOG # 137; \$55/hour) and Books in Print (DIALOG # 470, \$65/hour)**  
are familiar to most scholars in their printed versions. Automated searches are often very fast and, thus, comparatively cheap.
- **Philosopher's Index (DIALOG #57, \$55/hour)**  
is similar to its printed version. International and comprehensive in the philosophical literature and the history of ideas.
- **LCMARC (DIALOG # 426; \$45/hour).**  
An excellent computerized card catalogue. Unfortunately, it holds only relatively recent LC acquisitions and so is limited in its utility for this generation of historians. To make matters worse, foreign languages entries are even more shallow. An excellent means of updating older bibliographies. Cheap. All in all an enormously important research tool.
- **REMARC (DIALOG # 421-425; \$85/hour).**  
These files represent a commercial effort to supply the pre-1980 data that LCMARC has left out. Although they are considerably more expensive than LCMARC, their utility as bibliographic sources is obvious.

### **Databasing: A Selected Bibliography**

*Computer Phone Book.* A Complete Annotated Listing of Names and Numbers So You Can Go On Line with Over 400 Systems Across the Country (Bergenfield, N.J., 1983).

Conger, L. D.: Limiting to Major. Database, vol., VI, no. 3, Aug., 1983, pp. 72-75.

Conger, L.D.: The Friendly User - Review of the Year Online: 1983, in: Database (U.S.), vol. VII, no. 2. June 1984, pp. 89-92.

- Cook, P. R.: Electronic Encyclopedias, in: *Byte* vol. IX, no. 7, July 1984, pp. 151-164.
- DeBardeleben, M. L., C. West, L. Cook, M. Rosenberg: The Philip Morris Information Network. A Library Database on an In-House Timesharing System, in: *Database*, vol. VI, no. 2, June, 1983, pp. 40-51.
- Dolan, D. R.: Staying on Top of Database Documentation, in: *Database*, vol. VI, no. 3, Aug., 1983, pp. 84-86.
- Droessler, J. B. and J. M. Rholes: Online Services at the Reference Desk: DIALOG, RILN, and OCLC, in: *Online*, vol. VII, no. 6, Nov., 1983 pp. 79-86.
- Edwards, J.: Desktop's On Line: Buyer's Guide to On-line Information Services, in: *Desktop Computing*, vol. III, no. 5, May, 1983, pp. 16-17, 19ff.
- Falk, J. D.: In Search of History: Bibliographic Databases, in: *History Teacher*, vol. XV, 1981, pp. 523-544.
- Falk, J. D.: The New Technology for Research in European Women's History: 'On-line' Bibliography, in: *Signs. A Journal of Women, Culture, and Society*, vol. IX, no. 1, pp. 120-33.
- Falk, J. D.: The Historian Enters the Electronic Age: Bibliographical and Database Publishing, in: *Public History*, vol. IV, no. 2, 1982, pp. 35-42.
- Ferrarini, E.: Doing Research with an On-Line Library, in: *Business Computer Systems*, vol. II, no. 2, Feb, 1983, pp. 39-40, 42.
- Ferrarini, E.: The Latest From Information Utilities. *Business Computer Systems*, vol. II, no. 3, March, 1983, pp. 33-34.
- Glossbrenner, A.: The Complete Handbook of Personal Computer Communications: Everything You Need to Know to Go Online with the World, New York, 1983.
- Glossbrenner, A.: Looking it up on Knowledge Index, in: *PC (Independent Guide to IBM Personal Computers)*, vol. III, no. 14, July, 1984, pp. 369-372.
- Guide to DIALOG Searching*, Palo Alto, CA, 1984.
- Halperin, M. and R. A. Pagell: Searchers' Perceptions of Online Database Vendors, in: *Special Libraries*, vol. LXXIV, no. 2, April, 1983, pp. 119-126.
- Hewes, J. J.: DIALOG: the Ultimate On-Line Library, in: *PC World*, vol. I, no. 6, Sept., 1983, pp. 74-88.
- Hall, J. L. and M. J. Brown: *Online Bibliographic Databases: An International Directory*, London, 1981.
- Jagodzinski, A.: A Theoretical Basis for the Representation of On-Line Computer Systems to Naive Users, in: *International Journal of Man-Machine Studies*, vol. XVIII, no. 3, March 1983, pp. 215-252.
- Krajewski, R.: Database Types. A Brief Look at Different Ways of Managing Information, in: *Byte*, vol. IX, no. 11, October 1984, pp. 137-142.

- Lesko, M.*: The Computer Data and Database Source Book, New York, 1984.
- Lesko, M.*: Low-Cost On-Line Databases. Cheaper Ways to Obtain Information With your Computer, in: *Byte*, vol. IX, no. 11, October 1984, pp. 167-174.
- Markey, K.*: Offline and Online User Assistance for Online Catalog Searchers, in: *Online*, vol. 8, no. 3, May 1984, pp. 54-66.
- Meadows, L. L.*: On-Line Libraries, in: *PC (Independent Guide to IBM Personal Computers)*, vol. III, no. 11, June, 1984.
- Meilach, D. Z.*: DIALOG Training Seminar, in: *Interface Age*, vol. VIII, no. 6, June, 1983, pp. 159-161.
- Palmer, R. C.*: Online Reference and Information Retrieval, Littleton, CO, 1983.
- Raben J. and G. Marks, eds.*: Databases in the Humanities and Social Sciences, Amsterdam, 1980.
- Rockman, I. F.*: Money Online: How to Search the Grant Literature, in: *Database*, vol. VI, no. 2, June, 1983, pp. 26-33.
- Rogers, R. T.*: The Wide World of DIALOG and the Knowledge Index Alternative, in: *Lawyer's PC*, vol. I, no. 18, May, 1984, pp. 1-7.
- Roper, M.*: ADP in Archives: a Third Bibliography, in: *Journal of the Society of Archivists (U. K.)*, vol. VI, no. 8, 1981, pp. 524-528.
- Shapiro, N. L.*: The Small Computer Connection: Telecommunications for the Home and Office, New York, 1983.
- Siegel, Jay*: Moving Data Between PCs and Mainframes, in: *Byte*, vol. IX, no. 9, Fall, 1984, pp. 248-255.
- Sweetland, J. H.*: America: History and Life, a Wide Ranging Database, in: *Database*, vol. VI, no. 4, December, 1983.
- Ward, J.*: The Micro in the Media Center: Online Searching: The Future is now with 'Knowledge Index', in: *Educational Computer*, vol. III, no. 3, May/June, 1983, pp. 28-29.